2013

Characterization of acyl-ACP thioesterases for the purpose of diversifying fatty acid synthesis pathway

Fuyuan Jing
Iowa State University

Characterization of acyl-ACP thioesterases for the purpose of diversifying fatty acid synthesis pathway

by

Fuyuan Jing

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Biochemistry

Program of Study committee:
Basil J. Nikolau, Major Professor
Marna D. Yandeau-Nelson
Mark S. Hargrove
Eve Syrkin Wurtele
Jacqueline V. Shanks

Iowa State University
Ames, Iowa
2013

TABLE OF CONTENTS

ACKNOWLEDGEMENTS vii

ABSTRACT ix

CHAPTER 1 - INTRODUCTION 1
Fatty acid synthesis pathway 1
Isolation and characterization of acyl-ACP TE 3
Classification and evolution of acyl-ACP TE 4
Catalytic mechanism of acyl-ACP TE 6
Determination of the Substrate specificity of acyl-ACP TE 6
Application of acyl-ACP TE in the metabolic engineering 7
Objectives of this study 9
References 9
Figures 14

CHAPTER 2 – PHYLOGENETIC AND EXPERIMENTAL CHARACTERIZATION OF AN ACYL-ACP THIOESTERASE FAMILY REVEALS SIGNIFICANT DIVERSITY IN ENZYMATIC SPECIFICITY AND ACTIVITY 15
Abstract 15
Background 16
Experimental Procedures 18
Phylogenetic analyses 18
DNA synthesis 19
Cloning of acyl-ACP TE cDNAs from Coconut (*Cocos nucifera*) and *Cuphea viscosissima* 19
In vivo activity assay 21
Identification of methylketone 2-tridecanone 22

Statistical cluster analysis 22
Results 22
Phylogenetic analysis and identification of acyl-ACP TEs 23
Isolation and sequence analysis of acyl-ACP TEs from coconut and *C. viscosissima* 25
Determination of in vivo activities of acyl-ACP TEs 26
Clustering acyl-ACP TEs based on their catalytic functionality 28
Discussion 29
The systematic functional characterization of bacterial acyl-ACP TEs demonstrates production of SCFAs 29
Acyl-ACP TEs from MCFA-producing plant tissues make MCFAs 30
Acyl-ACP TEs can intercept both saturated and unsaturated intermediates of Type II fatty acid synthase of *E. coli* 31
Subtle changes in primary sequences may be sufficient to change the substrate specificity of acyl-ACP TEs 32
Unexpected activity reveals diversity of acyl-ACP TEs 33
Conclusions 34
Acknowledgments 34
Authors' contributions 35
References 35
Tables 40
Figures 44

CHAPTER 3 – IDENTIFICATION OF ACTIVE SITE RESIDUES IMPLIES A TWO-STEP CATALYTIC MECHANISM FOR ACYL-ACP THIOESTERASE 51
Abstract 51
Introduction 52
Methods 54
Computational modeling and predictions 54
Site-directed mutagenesis 54

In vivo activities of acyl-ACP TE mutants 55
Determination of acyl-ACP TE protein expression level 55
Purification of acyl-ACP TE mutant proteins 56
CD spectra of acyl-ACP TE proteins 57
Results 57
Sequence alignment analysis of acyl-ACP thioesterases 57
Predicted structure of a plant acyl-ACP TE 58
Mutagenesis-based evaluation of potential catalytic residues of acyl-ACP TE 59
Discussion 61
Cys348 is not a catalytic residue 62
Identification of Asp309 and Glu347 as catalytic residues 63
A proposed catalytic mechanism for acyl-ACP TEs 64
Acknowledgments 66
References 66
Figures 69
Supplemental data 75

CHAPTER 4 – DISSECTING THE STRUCTURAL DETERMINANTS OF CHAIN LENGTH SELECTIVITY OF PLANT ACYL-ACP THIOESTERASES 77
Abstract 77
Introduction 78
Methods 81
Construction of chimeric acyl-ACP TEs 81
Multiple sequence alignment of functionally characterized acyl-ACP TEs 81
Structural modeling of CvFatB2 82
Site-directed mutagenesis 82
In vivo activity of acyl-ACP TE variants 82
Statistical cluster analysis 83
Determination of acyl-ACP TE protein expression level 83

Results 84
Substrate specificity of chimeric acyl-ACP TEs 84
Predicted tertiary structures of CvFatB2 85
Sequence alignment of functionally characterized acyl-ACP TEs 86
Site-directed mutagenesis analysis 87
Insight into the structure of a bacterial acyl-ACP TE 90
Discussion 92
The region that determines substrate specificity 92
Residues determining substrate specificity 93
Residues that affect the catalytic activity of acyl-ACP TE 94
A proposed model for substrate specificity 96
Acknowledgement 97
References 97
Tables 100
Figures 102
Supplemental material 111

CHAPTER 5 – DIRECTED EVOLUTION OF ACYL-ACP THIOESTERASE FOR HIGHER FATTY ACID PRODUCTION 117
Abstract 117
Introduction 118
Methods 120
Design of the primers 120
Generation of the mutant library 121
Screening for TE activity on the Neutral Red Plates 122
Analysis of the fatty acid production with GC-MS 122
Results 123
Characterization of the variant acyl-ACP TE library 123
Screening for TE activity on Neutral Red plates 123
Quantification of the fatty acid productivity of TE variants 124

Screening TE variants for higher fatty acid productivity.....124
Sequences of the mutants.....125
The substrate specificities of TE mutants.....126
Discussion.....126
Acknowledgement.....128
References.....129
Tables.....132
Figures.....134

CHAPTER 6 – CONCLUSIONS.....141

ACKNOWLEDGEMENTS

I'm grateful to my advisor, Dr. Basil J. Nikolau, for providing me with continued support, encouragement, and valuable guidance.

I'm also grateful to Dr. Marna D. Yandeau-Nelson for valuable instructions and guidance on my research.

I would like to thank…

Drs. Mark S. Hargrove, Eve Syrkin Wurtele, and Jacqueline V. Shanks for serving on my Program of Study committee and giving me many good suggestions.

Our lab manager, Dr. Libuse Brachova, for assistance with a lot of general lab business.

Dr. M. Ann D.N. Perera and Dr. Zhihong Song of the W.M. Keck Metabolomics Research Laboratory at Iowa State University for assistance with fatty acid analysis.

Dr. David Cantu in Dr. Peter Reilly's group for performing the phylogenetic analysis of acyl-ACP TEs.

Paul Kapke and Amanda Brockman of Hybridoma Facility at the Iowa State University for making the CvFatB2 antibody

Joel Nott of the Protein Facility at the Iowa State University for assistance with the CD spectroscopy.

Jarmila Tvaruzkova for cloning the acyl-ACP TEs from *Cuphea viscossima*

Sara Pederson for helping with the molecular biology experiment and fatty acid extraction in directed evolution of acyl-ACP TE.

Colin Hueser for helping with site-directed mutagenesis of CvFatB2 and purification of those mutant proteins to verify the catalytic residues.

Derek Loneman and Sam Condon for helping with the fatty acid extraction

REU students, Sumira Stein and Emily Lin, for helping me with the experiments during the summers.

Shivani Garg, Bryon Upton, Lucas Showman, Xin Guan, Daolin Chen, and Huanan Jin for many useful conversations and advices.

All the members in Dr Nikoau's group for creating a friendly, helpful, and encouraging lab environment.

I want to thank my friends, colleagues, the department faculty and staff for making my time at Iowa State University a wonderful experience.

Finally, I would like to thank my girl friend Le Zhao for supporting my work and helping me conduct some key experiments. Great thanks to my families for supporting and encouraging my studying in the US.

ABSTRACT

Acyl-ACP TE selectively hydrolyzes the thiol ester bonds of acyl-ACPs to release free fatty acids, and therefore plays an essential role in determining the output of fatty acid synthesis (FAS) pathway. Comprehensive understanding of acyl-ACP TE is demanded to tailor this biocatalyst for the application in metabolic engineering of FAS pathway. To explore the diversity of acyl-ACP TEs, a total of 31 TEs enzymes were sourced from a wide range of biological taxa, including plants and bacteria, and these were functionally characterized. The results demonstrate that acyl-ACP TEs have great functional diversity relative to the acyl chain length specificity as well as acyl chains that contain additional chemical functionalities. Multiple sequence alignment of plant and bacterial TEs, and structure modeling of CvFatB2 revealed that a previously proposed residue Cys348 is unlikely to be a catalytic residue. Instead, residues Asp309 and Glu347, in addition to previously proposed residues Asn311 and His313 (numbers are based on CvFatB2 sequence), were proposed to be involved in the catalysis of acyl-ACP TEs. *In vivo* activities of site-directed mutants proved this hypothesis, and a two-step catalytic mechanism for plant and bacterial acyl-ACP TEs is proposed. To identify the region(s) that determine the substrate specificity, two acyl-ACP TEs were used for a domain-shuffling study. Comparing the substrate specificities of the resulting chimeric TEs led to the identification of the most important region that determines the substrate specificity of acyl-ACP TE. Site-directed mutagenesis analysis proved that six residues play critical roles in determining the substrate specificity, including V194 in Fragment II, V217, N223, R226, and R227 in Fragment III, and I268 in Fragment IV. Another three residues, L257, I260, and L289, impact the catalytic activity of acyl-ACP TE, because they are in two proposed ACP binding motifs. A directed evolution approach was successfully developed to improve the fatty acid productivity of acyl-ACP TE. Screening a designed variant library resulted in recovery of TE variants with increased fatty acid productivity and more insight into the relationship between sequences and substrate specificities of acyl-ACP TE.

CHAPTER 1 - INTRODUCTION

Fatty acids are very important molecules as structural, signaling, and energy-storage components in living systems. Naturally stored in the form of triacylglycerol in the plant seed oil or animal fat, fatty acids are the most abundant renewable resource of highly reduced carbon. They have been consumed by humans not only as food but also historically for many thousands of years as feedstocks for different non-food applications, most recently in oleochemical industry. Fatty acids are used to make ingredients of a lot of products, such as soaps, detergents, surfactants, lubricants, cosmetics, and pharmaceuticals (Ohlrogge 1994; Thelen et al. 2002; Dyer et al. 2008). Recently, the use of fatty acids for the production of high-energy density biofuels or biorenewable chemical feedstocks has been increasing, mostly driven by the rapidly rising cost of petroleum and growing environmental concerns about consuming large amount of fossil fuel (Durrett et al. 2008; Nikolau et al. 2008).

The fatty acid synthesis (FAS) pathway has drawn intensive attention as a route for the production of fatty acids. Acyl-acyl carrier protein (ACP) thioesterase (TE), representing the terminal step in the FAS pathway, plays crucial role in controlling the metabolic flux through the FAS pathway. It has also received significant interest in the metabolic engineering of FAS pathway. The background and objectives of my research is introduced in this chapter.

Fatty acid synthesis pathway

Living organisms can synthesize a variety of fatty acids *de novo* from the small precursor, acetyl-CoA. The first committed reaction of fatty acid biosynthesis is catalyzed by acetyl-CoA carboxylase (ACCase), which carboxylates acetyl-CoA to form malonyl-CoA. In the following reaction, the malonyl-CoA-ACP transferase (MAT) transfers the malonyl group from CoA to ACP, a ~8 KDa acidic protein which serves as acyl chain carrier molecule in the FAS pathway. The resulting malonyl-ACP is the elongation unit for the iterative fatty acyl chain elongation cycle. The acyl chain

elongation starts with the condensation of malonyl-ACP and acetyl-CoA catalyzed by a β-ketoacyl-ACP synthase III (KAS III) to generate acetoacetyl-ACP, followed by three consecutive reduction reactions catalyzed by β-ketoacyl-ACP reductase (KR), β-hydroxylacyl-ACP dehydrase (HD), and enoyl-ACP reductase (ER) to yield a butyl-ACP (Figure 1.1). The following condensation reactions of malonyl-ACP and acyl-ACPs are catalyzed by KAS I or KAS II enzymes. The condensation reduction cycle can repeat several times with the growth of the acyl chain length by two-carbon units per cycle. The elongation process usually ends at 16-18 carbon acyl-ACPs, which are incorporated into phospholipids by acyltransferases or hydrolyzed by acyl-ACP TE to release the free fatty acids.

There are two types of FAS architectures, FAS I and FAS II. Animals synthesize fatty acids with FAS I, which consists of multi-functional polypeptides with several domains each catalyzing the individual reactions of the FAS pathway (Smith et al. 2003). FAS II exists mainly in bacteria and plants, but also found in the mitochondria of eukaryotes (White et al. 2005; Hiltunen et al. 2009). Different from FAS I, FAS II consists of a series of individual enzymes that are encoded by discrete genes, each of which catalyzes the individual reactions of the FAS pathway. This study focused on FAS II, because the intermediates of FAS II are more accessible for metabolic engineering.

The regulation of fatty acid biosynthesis has been widely studied, and a lot of progress has been made. So far, the best understanding of the regulation comes from the study of *Escherichia coli* FAS pathway. The regulation of FAS in *E. coli* has been reviewed in several recent papers (Magnuson et al. 1993; Fujita et al. 2007; Lennen et al. 2012). Generally, the FAS pathway is tightly regulated at different levels, because it affects the fatty acids incorporated into the membrane and thus influences the membrane physiology. At the transcriptional level, two transcriptional regulators, FabR and FadR, are involved in the regulation of both fatty acid biosynthesis and β-oxidation (the degradation of fatty acids). The FAS pathway is also regulated at the enzymology level. In *E. coli*, the key regulatory signals that control the FAS pathway are long chain acyl-ACPs. Many studies have shown that long chain acyl-ACP can directly inhibit the

activity of ACCase, the first enzyme in the pathway (Jiang et al. 1994; Davis et al. 2001). The other two enzymes, β-ketoacyl-ACP synthase III (FabH) and enoyl-ACP reductase (FabI), may be also inhibited by long chain acyl-ACPs, but to a lesser extent (Heath et al. 1996). The in-depth understanding of the regulation of the FAS pathway is highly valuable for the metabolic engineering of this pathway for the production of fatty acids.

Isolation and characterization of acyl-ACP TE

Acyl-ACP TE is the enzyme that catalyzes the hydrolysis of the thioester bond of acyl-ACP to release free fatty acid and ACP. In the 1990s, the first medium chain-specific acyl-ACP TE was identified and purified from immature cotyledons of *Umbellularia californica* (California Bay) in an effort to understand the mechanism for the synthesis of medium chain fatty acids (about C-12 chain length) in the seeds of species that specifically accumulate those fatty acids (Davies et al. 1991; Pollard et al. 1991). Later, the cDNA of this 12:0-ACP TE was isolated and expressed in *Arabidopsis* and *Brassica napus*, resulting in significant accumulation of lauric acid in the transgenic seeds (Voelker et al. 1992). This result proved that acyl-ACP TEs play important roles in the chain length determination during fatty acid biosynthesis. Following the isolation of the first acyl-ACP TE, many studies focused on the isolation and characterization of acyl-ACP TEs from other plant species that accumulate medium chain fatty acids in their seed oils. Especially, many acyl-ACP TEs were cloned and characterized from species in the genus *Cuphea* (Dehesh et al. 1996; Dehesh et al. 1996; Leonard et al. 1997). For example, *Cuphea palustris*, *Cuphea hookeriana*, *Cuphea wrightii*, and *Cuphea lanceolata*, because their seed oils contain a variety of medium chain (8:0-14:0) fatty acids (Graham 1989). Most of the acyl-ACP TEs were functionally characterized by heterologous expression in *Arabidopsis*, *B. napus*, or *E. coli* (Voelker et al. 1992; Voelker et al. 1994). The significant change in the fatty acid profile demonstrated the important roles of acyl-ACP TEs in determining the composition of storage lipids.

It was also noticed that the fatty acid composition of the transgenic seeds was not exactly the same as that in the original species, suggesting that acyl-ACP TE may not be

the only enzyme that determines the fatty acid composition (Leonard et al. 1997). It was proposed that the kinetics of the entire FAS pathway may determine the fatty acid profile because acyl-ACP TEs compete with β-ketoacyl-ACP synthase for the acyl-ACP intermediates (Voelker et al. 1994; Leonard et al. 1997). A computational simulation was made to demonstrate that the increasing interception of shorter chain acyl-ACPs leads to the reduction of longer chain acyl-ACP pools and makes them less available for TE hydrolysis (Voelker et al. 1994). The regulatory role of KAS enzyme in the medium chain fatty acid synthesis was later emphasized with the identification and characterization of a KAS IV enzyme, which displayed strong preference for the elongation of short chain and medium chain acyl-ACPs (C4- to C10-ACP) but showed less activity for the further elongation reaction of long chain acyl-ACPs (Schutt et al. 2002). To summarize the mechanisms involved in medium chain fatty acid biosynthesis, the KAS enzymes determine the elongation rate of acyl-ACPs of different chain lengths and control the pool size of each acyl-ACP, while acyl-ACP TEs are capable of selectively intercepting the acyl-ACPs to produce specific fatty acids.

Most of the previous studies were focused on plant acyl-ACP TEs. The bacterial acyl-ACP TEs were rarely studied. The physiological role of acyl-ACP TEs has been well established, but the physiological role of bacterial acyl-ACP TEs is still not clear. In bacteria, the long chain acyl-ACPs produced by FAS pathway are directly used for the synthesis of phospholipids by acyltransferases (i.e., PlsB, PlsC, PlsX, and PlsY) (Zhang et al. 2008).

Classification and evolution of acyl-ACP TE

In plants, fatty acid biosynthesis occurs predominantly in plastids. Plant acyl-ACP TEs are nuclear-encoded plastid-targeted proteins (Voelker et al. 1992). The increasing number of plant acyl-ACP TEs allowed better analysis of their sequence structures. Based on multiple sequence alignment, plant acyl-ACP TEs were classified into two classes, FatA and FatB (Jones et al. 1995). High sequence similarity was found within each group as well as between groups. Both FatA and FatB have N-terminal

transit peptides, which are about 60-90 amino acids (Jones et al. 1995; Ghosh et al. 2007; Sanchez-Garcia et al. 2010). The N-terminal sequences of the mature proteins of FatB TEs are conserved as LPDWS. One characteristic feature of FatB TEs is the presence of a highly hydrophobic region at the N-terminus of the mature protein, which is not present in the FatA TEs. This region was predicted as a membrane anchor and the removal of this region did not affect the activity or specificity of acyl-ACP TE (Facciotti et al. 1998). The FatA sequences differ from FatB sequences by two insertions in the C-terminal half of the protein (Jones et al. 1995).

FatA and FatB TEs not only have different sequences, but also have distinct substrate specificities. FatA TEs consistently hydrolyze 18:1-ACP, while FatB TEs prefer saturated acyl-ACPs with acyl chain length varying from 8- to 18-carbons. Additionally, FatB TEs fall into two groups: one group is active on C14 to C18 acyl-ACPs with predominant preference for 16:0-ACP and is expressed throughout the plant; and another group is specific for saturated acyl-ACPs in the C8-C14 range and its expression is restricted to the tissues that accumulate medium chain fatty acids (Jones et al. 1995; Dehesh et al. 1996).

The coexistence of different types of acyl-ACP TEs raises an interesting question: what is the evolutionary relationship between those TEs? Based on the phylogenetic analysis of plant acyl-ACP TEs, it was proposed that the 16:0-ACP specific FatB is the ancient form and it's ubiquitous in plants. The C8 to C14-ACP specific FatB and the ubiquitous 18:1-ACP specific FatA were evolved from this 16:0-ACP TE (Jones et al. 1995).

To date, most of the functionally characterized acyl-ACP TEs are from higher plants. In the past decade, the annotated acyl-ACP TE sequences in the public database have increased exponentially, including a large number of bacterial acyl-ACP TEs that have never been functionally characterized. Study of the acyl-ACP TEs in bacteria or lower plants will help us better understand the classification and evolution of this enzyme.

Catalytic mechanism of acyl-ACP TE

Understanding the catalytic mechanism of acyl-ACP TE will better allow the engineering of better biocatalysts. A lot of efforts have been attempted to identify the catalytic residues of plant acyl-ACP TEs. The initial evidences came from enzymatic assays with different inhibitors. Plant acyl-ACP TEs showed sensitivity to thiol inhibitors but not to serine hydroxyl-reactive reagents, suggesting a cysteine was involved in catalysis (Davies et al. 1991; Pollard et al. 1991). Mutating a conserved cysteine to alanine completely inactivated the enzyme, while mutating the cysteine to serine resulted in an enzyme that retained ~60% activity and converted inhibitor sensitivity from thiol reagents to serine reagents (Yuan et al. 1996). When treated with diethylpyrocarbonate, plant acyl-ACP TE lost 97% of its activity, suggesting the existence of an active-site histidine (Davies et al. 1991). This catalytic histidine was later identified by site-directed mutagenesis (Yuan et al. 1996). Based on a predicted structural model of plant acyl-ACP TE and site-directed mutagenesis, a specific residue Asn was identified as the third catalytic residue (Mayer et al. 2005). Based on these results, plant acyl-ACP TEs have been proposed to utilize a papain-like catalytic triad containing Asn, His, and Cys (Yuan et al. 1996). However, without higher order structural information and kinetic characteristic data, the catalytic sites of acyl-ACP TEs have yet to be completely proven and the catalytic mechanism is still obscure.

Plant and bacterial acyl-ACP TEs have been proposed to have the same hotdog structure, though they share very little sequence similarity (Cantu et al. 2010). As more and more plants as well as bacterial acyl-ACP TEs have been functionally characterized and some bacterial TEs structures have been solved (i.e., PDB structure 2OWN, 2ESS, and 4GAK), it is worth to reinvestigate the catalytic residues of acyl-ACP TEs taken into account this more extensive structural information.

Determination of the Substrate specificity of acyl-ACP TE

FatB TEs share high amino acid sequence identity, but display different substrate specificities. It is speculated that the substrate specificity may be determined by a small

set of residues (Jones et al. 1995). Understanding the structural basis for the determination of substrate specificity is crucial for engineering acyl-ACP TEs of desired selectivity. Because there is no crystal structure for plant acyl-ACP TEs, previous attempts to find amino acids involved in substrate specificity determination mainly relied on domain shuffling between different acyl-ACP TEs, and analyzing the substrate specificity of the resulting chimeric enzymes. By comparing two chimeric enzymes generated from a 12:0-ACP TE (UcFatB1) from *U. californica* and a 14:0-ACP TE (CcFatB1) from *Cinnamonum camphorum*, it was concluded that the C-terminal two-thirds of the protein determines the substrate specificity (Yuan et al. 1995). A double mutant (M197R/R199H) within this region turned UcFatB1 from 12:0-ACP specific to 12:0/14:0-ACP specific with equal preference for both substrates (Yuan et al. 1995). In another study, characterization of chimeric enzymes generated from *Arabidopsis* acyl-ACP TE AtFatA and AtFatB demonstrated that the N-terminus of the protein is critical for determination of substrate specificity (Salas et al. 2002). Structural modeling of *Arabidopsis* AtFatB revealed that plant acyl-ACP TEs adopt hotdog structure and the N-terminal domain contains residues that affect substrate specificity (Mayer et al. 2005; Mayer et al. 2007). Although some success has been achieved in identifying determinant amino acid residues for substrate specificity, the complete structural basis for specificity is still obscure.

Application of acyl-ACP TE in the metabolic engineering

Initially, the motivation for studying acyl-ACP TEs is to understand the mechanism for medium chain fatty acid synthesis and use that mechanism to modify the fatty acid composition in domestic oil crops, because medium chain fatty acids, having many industrial applications, are not normally produced by US domestic oil crops. For example, expression of a 12:0-ACP TE in *Arabidopsis* and *B. napus* resulted in significant accumulation of lauric acid in the seed oil (Voelker et al. 1992; Voelker 1996). Overexpression of ChFatB2, a medium chain specific TE, in *B. napus* also lead to

substantial accumulation of 8:0 and 10:0 fatty acids in the transgenic seeds (Dehesh et al. 1996).

Recently, acyl-ACP TE, as a critical element in the metabolic engineering of FAS pathway, has attracted even more attention because of increasing interest in producing fatty acids as biofuel or biorenewable chemicals. A lot of efforts have been made to engineer the FAS pathway in microbes or algae for the production of free fatty acids or fatty acid derivatives (Liu et al. 2011; Zhang et al. 2011; Ranganathan et al. 2012; Zhang et al. 2012). Among all those efforts focused on improving free fatty acid production in microbes, mainly *Escherichia coli*, two approaches have been demonstrated to be the most critical steps (Lennen et al. 2012). One is the overexpression of acyl-ACP TE to release fatty acids from the FAS pathway, and another one is impairing the fatty acid β-oxidation pathway with fadD or fadE mutations to prevent the reuse of free fatty acids. Acyl-ACP TEs not only control the fatty acid composition by selectively hydrolyzing the acyl-ACPs, but also enhance the fatty acid production through two effects: (1) they create a product sink; (2) they elevate the metabolic flux through the FAS pathway by depleting long chain acyl-ACPs, the feedback inhibitors for upstream enzymes in the FAS pathway (Jiang et al. 1994; Heath et al. 1996; Lennen et al. 2012). Optimization of the expression level of acyl-ACP TEs with different plasmid copy numbers or different promoters leads to significant improvement of free fatty acid production in *E coli*, also demonstrating the crucial role of acyl-ACP TE in metabolic engineering of the FAS pathway for the production of free fatty acids (Steen et al. 2010; Youngquist et al. 2012; Zhang et al. 2012). All the acyl-ACP TEs used in the previous studies are from natural resources, which have evolved for the organisms to survive in different environments but may not be optimal for industry applications, in terms of substrate specificity or catalytic efficiency. It is tempting to increase activity of acyl-ACP TE to help the improvement of fatty acid production in microbes.

Objectives of this study

Acyl-ACP TE controls the output of FAS machinery. To produce fatty acids of different chain lengths, it is essential to have acyl-ACP TEs with specific substrate specificity. Acyl-ACP TEs have displayed selectivity on chain length of the substrate. It is also possible that acyl-ACP TE can hydrolyze the acyl chain when it is not fully reduced, i.e., keto, hydroxyl, and unsaturated fatty acids, which are highly valuable in the chemical industry. Natural acyl-ACP TEs normally do not have the optimum catalytic efficiency and substrate specificity for industry application. So, the objective of this study is to generate comprehensive knowledge about acyl-ACP TE so that we can tailor the biocatalyst for metabolic engineering of the FAS pathway. In this study, I first explored the acyl-ACP TEs in nature, including synthesized TEs based on the sequences in the public database and TEs cloned from species that produce significant amounts of medium chain and short chain fatty acids. Then, I tried to identify the catalytic residues and catalytic mechanism of acyl-ACP TEs. I also elucidated the structural basis for the determination of substrate specificity of acyl-ACP TE. Lastly, a directed evolution approach of acyl-ACP TE was developed for screening higher fatty acid production.

References

Cantu, D. C., Chen, Y. and Reilly, P. J. (2010). Thioesterases: a new perspective based on their primary and tertiary structures. Protein Science 19(7): 1281-1295.

Davies, H. M., Anderson, L., Fan, C. and Hawkins, D. J. (1991). Developmental induction, purification, and further characterization of 12:0-ACP thioesterase from immature cotyledons of Umbellularia californica. Archives of Biochemistry and Biophysics 290(1): 37-45.

Davis, M. S. and Cronan, J. E. (2001). Inhibition of Escherichia coli acetyl coenzyme A carboxylase by acyl-acyl carrier protein. Journal of Bacteriology 183(4): 1499-1503.

Dehesh, K., Edwards, P., Hayes, T., Cranmer, A. M. and Fillatti, J. (1996). Two novel thioesterases are key determinants of the bimodal distribution of acyl chain length of Cuphea palustris seed oil. Plant Physiology 110(1): 203-210.

Dehesh, K., Jones, A., Knutzon, D. S. and Voelker, T. A. (1996). Production of high levels of 8:0 and 10:0 fatty acids in transgenic canola by overexpression of Ch FatB2, a thioesterase cDNA from Cuphea hookeriana. Plant Journal 9(2): 167-172.

Durrett, T. P., Benning, C. and Ohlrogge, J. (2008). Plant triacylglycerols as feedstocks for the production of biofuels. Plant Journal 54(4): 593-607.

Dyer, J. M., Stymne, S., Green, A. G. and Carlsson, A. S. (2008). High-value oils from plants. Plant Journal 54(4): 640-655.

Facciotti, M. T. and Yuan, L. (1998). Molecular dissection of the plant acyl-acyl carrier protein thioesterases. Fett-Lipid 100(4-5): 167-172.

Fujita, Y., Matsuoka, H. and Hirooka, K. (2007). Regulation of fatty acid metabolism in bacteria. Molecular Microbiology 66(4): 829-839.

Ghosh, S. K., Bhattacharjee, A., Jha, J. K., Mondal, A. K., Maiti, M. K., Basu, A., Ghosh, D., Ghosh, S. and Sen, S. K. (2007). Characterization and cloning of a stearoyl/oleoyl specific fatty acyl-acyl carrier protein thioesterase from the seeds of Madhuca longifolia (latifolia). Plant Physiology and Biochemistry 45(12): 887-897.

Graham, S. A. (1989). Cuphea - a New Plant Source of Medium-Chain Fatty-Acids. Critical Reviews in Food Science and Nutrition 28(2): 139-173.

Heath, R. J. and Rock, C. O. (1996). Inhibition of beta-ketoacyl-acyl carrier protein synthase III (FabH) by acyl-acyl carrier protein in Escherichia coli. Journal of Biological Chemistry 271(18): 10996-11000.

Heath, R. J. and Rock, C. O. (1996). Regulation of fatty acid elongation and initiation by acyl acyl carrier protein in Escherichia coli. Journal of Biological Chemistry 271(4): 1833-1836.

Hiltunen, J. K., Schonauer, M. S., Autio, K. J., Mittelmeier, T. M., Kastaniotis, A. J. and Dieckmann, C. L. (2009). Mitochondrial Fatty Acid Synthesis Type II: More than Just Fatty Acids. Journal of Biological Chemistry 284(14): 9011-9015.

Jiang, P. and Cronan, J. E. (1994). Inhibition of Fatty-Acid Synthesis in Escherichia-Coli in the Absence of Phospholipid-Synthesis and Release of Inhibition by Thioesterase Action. Journal of Bacteriology 176(10): 2814-2821.

Jones, A., Davies, H. M. and Voelker, T. A. (1995). Palmitoyl-acyl carrier protein (ACP) thioesterase and the evolutionary origin of plant acyl-ACP thioesterases. Plant Cell 7(3): 359-371.

Lennen, R. M. and Pfleger, B. F. (2012). Engineering Escherichia coli to synthesize free fatty acids. Trends in Biotechnology 30(12): 659-667.

Leonard, J. M., Slabaugh, M. B. and Knapp, S. J. (1997). Cuphea wrightii thioesterases have unexpected broad specificities on saturated fatty acids. Plant Molecular Biology 34(4): 669-679.

Liu, X., Sheng, J. and Curtiss, R., 3rd (2011). Fatty acid production in genetically modified cyanobacteria. Proceedings of the National Academy of Sciences of the United States of America 108(17): 6899-6904.

Magnuson, K., Jackowski, S., Rock, C. O. and Cronan, J. E., Jr. (1993). Regulation of fatty acid biosynthesis in Escherichia coli. Microbiological Reviews 57(3): 522-542.

Mayer, K. M. and Shanklin, J. (2005). A structural model of the plant acyl-acyl carrier protein thioesterase FatB comprises two helix/4-stranded sheet domains, the N-terminal domain containing residues that affect specificity and the C-terminal domain containing catalytic residues. Journal of Biological Chemistry 280(5): 3621-3627.

Mayer, K. M. and Shanklin, J. (2007). Identification of amino acid residues involved in substrate specificity of plant acyl-ACP thioesterases using a bioinformatics-guided approach. BMC Plant Biology 7: 1.

Nikolau, B. J., Perera, M. A., Brachova, L. and Shanks, B. (2008). Platform biochemicals for a biorenewable chemical industry. Plant Journal 54(4): 536-545.

Ohlrogge, J. B. (1994). Design of New Plant Products: Engineering of Fatty Acid Metabolism. Plant Physiology 104(3): 821-826.

Pollard, M. R., Anderson, L., Fan, C., Hawkins, D. J. and Davies, H. M. (1991). A Specific Acyl-Acp Thioesterase Implicated in Medium-Chain Fatty-Acid Production in Immature Cotyledons of Umbellularia-Californica. Archives of Biochemistry and Biophysics 284(2): 306-312.

Ranganathan, S., Tee, T. W., Chowdhury, A., Zomorrodi, A. R., Yoon, J. M., Fu, Y., Shanks, J. V. and Maranas, C. D. (2012). An integrated computational and experimental study for overproducing fatty acids in Escherichia coli. Metabolic Engineering 14(6): 687-704.

Salas, J. J. and Ohlrogge, J. B. (2002). Characterization of substrate specificity of plant FatA and FatB acyl-ACP thioesterases. Archives of Biochemistry and Biophysics 403(1): 25-34.

Sanchez-Garcia, A., Moreno-Perez, A. J., Muro-Pastor, A. M., Salas, J. J., Garces, R. and Martinez-Force, E. (2010). Acyl-ACP thioesterases from castor (Ricinus

communis L.): an enzymatic system appropriate for high rates of oil synthesis and accumulation. Phytochemistry 71(8-9): 860-869.

Schutt, B. S., Abbadi, A., Loddenkotter, B., Brummel, M. and Spener, F. (2002). Beta-ketoacyl-acyl carrier protein synthase IV: a key enzyme for regulation of medium-chain fatty acid synthesis in Cuphea lanceolata seeds. Planta 215(5): 847-854.

Smith, S., Witkowski, A. and Joshi, A. K. (2003). Structural and functional organization of the animal fatty acid synthase. Progress in Lipid Research 42(4): 289-317.

Steen, E. J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., Del Cardayre, S. B. and Keasling, J. D. (2010). Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. Nature 463(7280): 559-562.

Thelen, J. J. and Ohlrogge, J. B. (2002). Metabolic engineering of fatty acid biosynthesis in plants. Metabolic Engineering 4(1): 12-21.

Voelker, T. A. (1996). Genetic engineering of a quantitative trait: Metabolic and genetic parameters influencing the accumulation of laurate in rapeseed (vol 9, pg 229, 1996). Plant Journal 9(5): 775-775.

Voelker, T. A. and Davies, H. M. (1994). Alteration of the specificity and regulation of fatty acid synthesis of Escherichia coli by expression of a plant medium-chain acyl-acyl carrier protein thioesterase. Journal of Bacteriology 176(23): 7320-7327.

Voelker, T. A., Worrell, A. C., Anderson, L., Bleibaum, J., Fan, C., Hawkins, D. J., Radke, S. E. and Davies, H. M. (1992). Fatty acid biosynthesis redirected to medium chains in transgenic oilseed plants. Science 257(5066): 72-74.

White, S. W., Zheng, J., Zhang, Y. M. and Rock (2005). The structural biology of type II fatty acid biosynthesis. Annual Review of Biochemistry 74: 791-831.

Youngquist, J. T., Lennen, R. M., Ranatunga, D. R., Bothfeld, W. H., Marner, W. D. and Pfleger, B. F. (2012). Kinetic modeling of free fatty acid production in Escherichia coli based on continuous cultivation of a plasmid free strain. Biotechnology and Bioengineering 109(6): 1518-1527.

Yuan, L., Nelson, B. A. and Caryl, G. (1996). The catalytic cysteine and histidine in the plant acyl-acyl carrier protein thioesterases. Journal of Biological Chemistry 271(7): 3417-3419.

Yuan, L., Voelker, T. A. and Hawkins, D. J. (1995). Modification of the substrate specificity of an acyl-acyl carrier protein thioesterase by protein engineering. Proceedings of the National Academy of Sciences of the United States of America 92(23): 10639-10643.

Zhang, F., Carothers, J. M. and Keasling, J. D. (2012). Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. Nature Biotechnology 30(4): 354-359.

Zhang, X. J., Li, M., Agrawal, A. and San, K. Y. (2011). Efficient free fatty acid production in Escherichia coli using plant acyl-ACP thioesterases. Metabolic Engineering 13(6): 713-722.

Zhang, Y. M. and Rock, C. O. (2008). Thematic review series: Glycerolipids - Acyltransferases in bacterial glycerophospholipid synthesis. Journal of Lipid Research 49(9): 1867-1874.

Figures

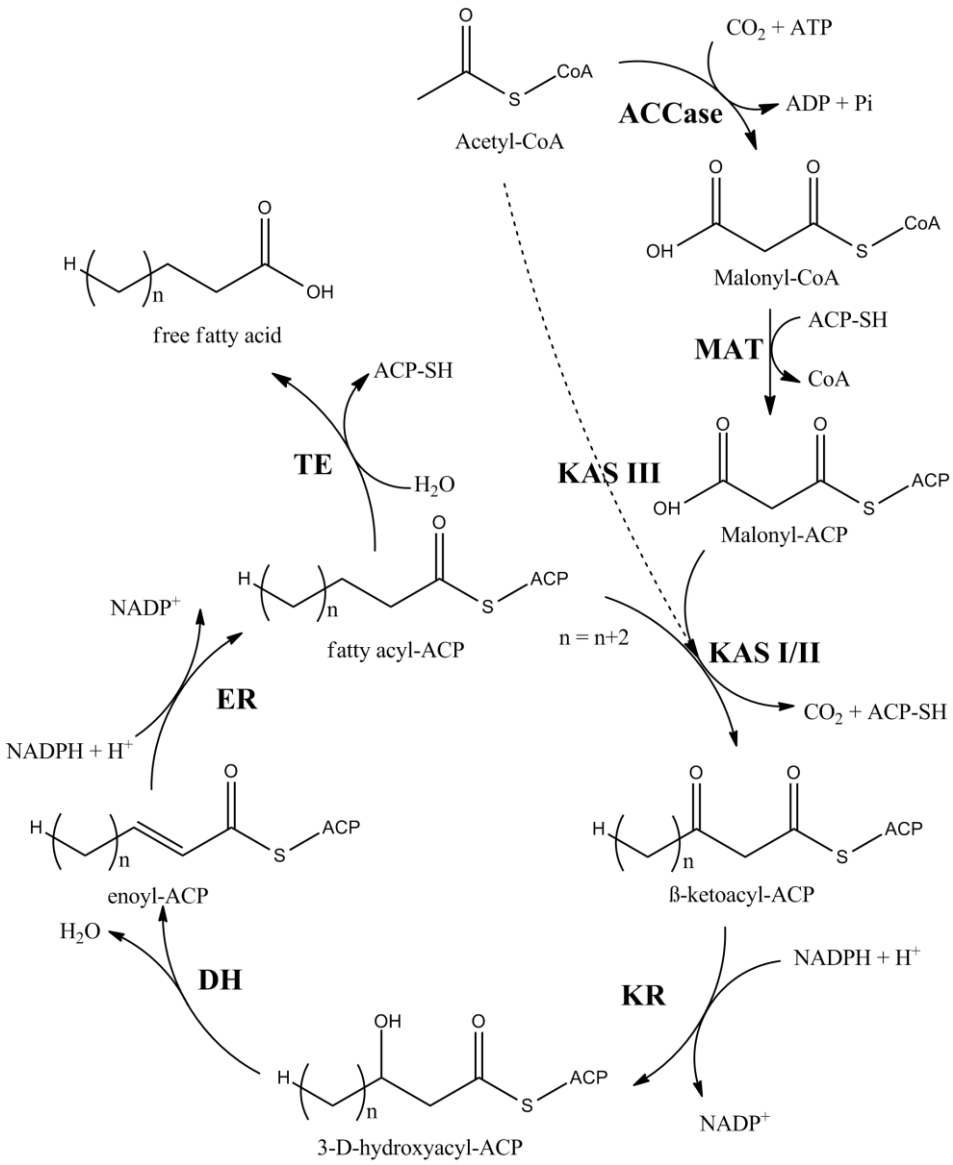


Figure 1.1 Fatty acid synthesis pathway

CHAPTER 2 – PHYLOGENETIC AND EXPERIMENTAL CHARACTERIZATION OF AN ACYL-ACP THIOESTERASE FAMILY REVEALS SIGNIFICANT DIVERSITY IN ENZYMATIC SPECIFICITY AND ACTIVITY

Fuyuan Jing[1], David C. Cantu[2], Jarmila Tvaruzkova[1], Jay P. Chipman[2], Basil J. Nikolau[1], Marna D. Yandeau-Nelson[1], and Peter J. Reilly[2]

[1]Department of Biochemistry, Biophysics, and Molecular Biology and [2]Department of Chemical and Biological Engineering, Iowa State University

Modified from a paper published on BMC Biochemistry with the same title and authors

Authors' contributions are listed at the end of this chapter

Abstract

Acyl-acyl carrier protein thioesterases (acyl-ACP TEs) catalyze the hydrolysis of the thioester bond that links the acyl chain to the sulfhydryl of the phosphopantetheine prosthetic group of ACP. This reaction terminates acyl chain elongation of fatty acid biosynthesis, and in plant seeds it is the biochemical determinant of the fatty acid compositions of storage lipids. To explore acyl-ACP TE diversity and to identify novel acyl ACP-TEs, 31 acyl-ACP TEs from wide-ranging phylogenetic sources were characterized to ascertain their *in vivo* activities and substrate specificities. These acyl-ACP TEs were chosen by two different approaches: 1) 24 TEs were selected from public databases on the basis of phylogenetic analysis and fatty acid profile knowledge of their source organisms; and 2) seven TEs were molecularly cloned from oil palm (*Elaeis guineensis*), coconut (*Cocos nucifera*) and *Cuphea viscosissima*, organisms that produce medium-chain and short-chain fatty acids in their seeds. The *in vivo* substrate specificities of the acyl-ACP TEs were determined in *E. coli*. Based on their specificities, these enzymes were clustered into three classes: 1) Class I acyl-ACP TEs act primarily on 14- and 16-carbon acyl-ACP substrates; 2) Class II acyl-ACP TEs have broad substrate specificities, with major activities toward 8- and 14-carbon acyl-ACP substrates; and 3) Class III acyl-ACP TEs act predominantly on 8-carbon acyl-ACPs.

Several novel acyl-ACP TEs act on short-chain and unsaturated acyl-ACP or 3-ketoacyl-ACP substrates, indicating the diversity of enzymatic specificity in this enzyme family. These acyl-ACP TEs can potentially be used to diversify the fatty acid biosynthesis pathway to produce novel fatty acids.

Background

De novo fatty acid biosynthesis can be considered an iterative "polymerization" process, commonly primed with the acetyl moiety from acetyl-CoA and with iterative chain extension occurring by reaction with malonyl-ACP. In most organisms this process optimally produces 16- and 18-carbon (C16 and C18) fatty acids. The enzyme that determines fatty acid chain length is acyl-acyl carrier protein thioesterase (acyl-ACP TE). This enzyme catalyzes the terminal reaction of fatty acid biosynthesis, acyl-ACP thioester bond hydrolysis to release a free fatty acid and ACP.

In discrete phyla and/or tissues of specific organisms (primarily higher plant seeds), thioester hydrolysis optimally produces medium-chain (C8–C14) fatty acids (MCFAs), which have wide industrial applications (e.g., producing detergents, lubricants, cosmetics, and pharmaceuticals) (Dehesh et al. 1996). TEs that specifically hydrolyze medium-chain acyl-ACP substrates have been studied widely (Voelker et al. 1992; Yuan et al. 1995; Dehesh et al. 1996). Short-chain fatty acids (SCFAs; e.g. butanoic acid and hexanoic acid) have more recently gained importance as potential biorenewable chemicals that could be derived from the fatty acid biosynthesis pathway (Nikolau et al. 2008). As a critical acyl chain termination enzyme, acyl-ACP TEs with desired substrate specificities are therefore important for engineering this pathway.

To date, dozens of acyl-ACP TEs have been functionally characterized and sorted into two classes, FatA and FatB (Jones et al. 1995). FatA-class TEs act on long-chain acyl-ACPs, preferentially on oleoyl-ACP (Jones et al. 1995; Hawkins et al. 1998; Serrano-Vega et al. 2005; Sanchez-Garcia et al. 2010), while FatB-class TEs preferably hydrolyze acyl-ACPs with saturated fatty acyl chains (Jones et al. 1995). The archetypical FatB-class TE was isolated from the developing seeds of California bay

(*Umbellularia californica*). This enzyme is specific for 12:0-ACP, and it plays a critical role in MCFA production (Pollard et al. 1991; Voelker et al. 1992). This discovery spurred isolation of additional MCFA-specific TEs from *Cuphea* (Dehesh et al. 1996; Dehesh et al. 1996; Leonard et al. 1997), *Arabidopsis thaliana* (Dormann et al. 1995), *Myristica fragrans* (nutmeg) (Voelker et al. 1997), and *Ulmus americana* (elm) (Voelker et al. 1997).

Recently, TEs obtained from public databases were classified into 23 families based on sequence and three-dimensional structure similarity (Cantu et al. 2010). These TEs were defined as enzymes that can hydrolyze any thioester bond irrespective of the chemical nature of the carboxylic acid and thiol molecules that constitute the substrates of these enzymes. The TE sequences are collected in the constantly updated ThYme database (Cantu et al. 2011). Of these 23 families, Family TE14 contains plant and bacterial acyl-ACP TEs involved in Type II fatty acid synthesis, whose reactions are catalyzed by discrete monofunctional enzymes. When this study was conducted (summer and fall 2010), Family TE14 contained 360 unique sequences, but only ~7% of these sequences, all of which were FatA and FatB TEs from higher plants, had been functionally characterized. The remaining ~220 bacterial acyl-ACP TEs were mostly generated from genomic sequencing projects and had never been functionally characterized.

Here we report the results of a two-pronged approach to identify acyl-ACP TEs with novel substrate specificities, which potentially could allow researchers to better infer biochemical properties of closely related sequences. This strategy includes the functional characterization of diverse acyl-ACP TEs 1) rationally chosen based on phylogenetic classification of the enzymes and 2) isolated from organisms that are known to produce MCFAs and SCFAs. Functional characterization of 31 acyl-ACP TEs from diverse organisms led to the discovery that several novel TEs can be used to produce short-chain and unsaturated fatty acids as well as methylketones.

Experimental Procedures

Phylogenetic analyses

Sequences from Family TE14 (Cantu et al. 2010) in the ThYme database were downloaded from the GenBank (Benson et al. 2011) and UniProt (Consortium 2010) databases. Fragments and incomplete sequences were removed, yielding 360 acyl-ACP TE sequences. A multiple sequence alignment (MSA) was generated from catalytic domains of these sequences using MUSCLE 3.6 (Edgar 2004) with default parameters. An unrooted phylogenetic tree based on the MSA was built using Molecular Evolutionary Genetics Analysis 4 (MEGA4) (Tamura et al. 2007). The minimum evolution algorithm was used due to its high effectiveness with large data sets (Desper et al. 2002), gaps were subjected to pairwise deletion, and an amino acid Jones-Taylor-Thornton (JTT) (Jones et al. 1992) distance model was chosen. The phylogenetic tree was further verified by a bootstrap test with 1000 replicates. The bootstrapped consensus tree was qualitatively analyzed and broken into apparent subfamilies. Statistical analysis was conducted to show that all sequences within a subfamily were more closely related to each other than to sequences in other subfamilies. Based on the MSA, JTT distances between all sequences were calculated and arranged into a *j* x *j* matrix, where *j* is the total number of sequences. Inter-subfamily distances and variances were determined using this matrix. For each apparent subfamily, a smaller *k* x *k* matrix, where *k* is the number of sequences in a given subfamily, was calculated. From this, intra-subfamily mean distances and variances were determined. These values were applied to the following equation to determine *z*:

$$z = \frac{\bar{x}_{ij} - \left(\bar{x}_{ii} + \bar{x}_{jj}\right)/2}{\sqrt{\frac{\sigma_i^2}{n_{ij}} + \frac{\sigma_{ii}^2}{n_{ii}} + \frac{\sigma_{jj}^2}{n_{jj}}}}$$

where $\bar{x}_{ij}$, $\bar{x}_{ii}$, and $\bar{x}_{jj}$ are the inter- and intra-subfamily mean distances, n_{ij}, n_{ii}, and n_{jj} are the total number of taxa used for each $\bar{x}$ value, and σ_i^2, σ_{ii}^2, and σ_{jj}^2 are the pooled inter- and intra-subfamily variances (Mertz et al. 2005).

A *z*-value >3.3 between two subfamilies shows that the difference between them is statistically significant to $p < 0.001$. If a *z*-value between two apparent subfamilies were <3.3, alternative apparent subfamilies were chosen and/or individual sequences were removed, and the statistical calculations were repeated. Subfamilies were finally defined with a phylogenetic tree in which all *z*-values exceeded 3.3, sometimes leaving some sequences outside any subfamily (i.e. non-grouped sequences).

All sequences within individual subfamilies were aligned using MUSCLE 3.6, and rooted phylogenetic trees were built in MEGA4 with the same tree and bootstrap parameters as described above. A few sequences from another subfamily (that with the highest *z*-value) were chosen to root individual subfamily trees.

DNA synthesis

cDNA sequences encoding acyl-ACP TEs were codon-optimized for expression in *E. coli* using the OptimumGene codon optimization program provided by GenScript USA (Piscataway, NJ, USA). Sequences were both synthesized and cloned into vectors by GenScript. *Bam*HI and *Eco*RI restriction sites were added to the 5' and 3' ends of each sequence, and products were cloned into the pUC57 vector.

Cloning of acyl-ACP TE cDNAs from Coconut (Cocos nucifera) and Cuphea viscosissima

Coconut fruits of different developmental stages were obtained from the USDA-ARS-SHRS National Germplasm Repository (Miami, FL, USA). Seeds of *C. viscosissima* were obtained from the North Central Regional Plant Introduction Station (NCRPIS, Ames, IA, USA). They were treated overnight with 0.1 mM gibberellic acid and then germinated in a growth chamber (Environmental Growth Chambers, Chagrin

Falls, OH) with 12 h of illumination at 25 ℃ followed by 12 h of darkness at 15 ℃. Seedlings were transplanted into soil and cultivated at NCRPIS. Seeds at different developmental stages were collected and flash-frozen in liquid nitrogen.

Acyl-ACP TE cDNAs were cloned from *C. viscosissima* and coconut via a homologous cloning strategy. MSAs of plant TE14 sequences revealed two conserved regions (RYPTWGD and NQHVNNVK), from which two degenerate primers, DP-F3 (5'-AGNTAYCCNACNT GGGGNGA-3') and DP-R3 (5'-TACTTNACRTTRTTNACRTGYTGRTT-3'), were designed. RNA was extracted from endosperm of nearly mature coconuts and immature seeds of *C. viscosissima* using the total RNA (plant) kit (IBI Scientific, Peosta, IA, USA). RNA was reverse-transcribed to cDNA using the SuperScript™ first-strand synthesis system for RT-PCR kit (Invitrogen, Carlsberg, CA, USA). PCR was performed in a 50-µL reaction mixture containing 20 ng cDNA, 1x *Pfx* buffer, 1 mM $MgSO_4$, 0.3 mM dNTP, 5.12 M DP-F3 and DP-R3 primers, and 0.5 U *Pfx* polymerase (Invitrogen) using a cycling program of 94 ℃ for 4 min, 35 cycles of 94 ℃ for 30 s, 52 ℃ for 30 s and 72 ℃ for 45 s, and a final extension step of 72 ℃ for 5 min. The expected ~350-bp products were identified by agarose gel electrophoresis, and their DNA bands were recovered using the QiaQuick gel extraction kit (Qiagen, Valencia, CA, USA) and cloned into the pENTR TOPO TA vector (Invitrogen). Using primers designed from the sequences of the cloned 350-bp fragments, the 5'- and 3'-ends of the cDNAs were obtained using the SMARTer RACE (rapid amplification of the cDNA ends) cDNA amplification kit (Takara Bio, Otsu, Japan).

For each acyl-ACP TE sequence, the full-length cDNA, minus the N-terminal chloroplast transit peptide, was amplified by PCR with primers engineered to introduce *Bam*HI and *Eco*RI restriction sites at the 5'- and 3'-ends, respectively. The PCR-amplified products were digested with *Bam*HI and *Eco*RI and cloned into the corresponding restriction sites of the pUC57 vector, which placed the acyl-ACP TE sequence under the transcriptional control of the *lac*Z promoter. The sequence of each construct was confirmed by sequencing both strands. Confirmed expression vectors of

coconut genes were transformed into *E. coli* strain K27, while sequences of *C. viscosissima* acyl-ACP TEs were synthesized after being codon-optimized.

In vivo activity assay

E. coli strain K27 contains a mutation in the *fadD* gene impairing β-oxidation of fatty acids, which results in the accumulation of free fatty acids in the growth medium (Overath et al. 1969; Klein et al. 1971). Each TE was expressed in *E. coli* K27, and free fatty acids that accumulated in the medium were extracted and analyzed. Four colonies for each construct were independently cultured in 2 mL LB medium supplemented with 100 mg/L carbicillin in 17-mL culture tubes. When the culture reached an OD_{600} of ~0.7, the growth medium was replaced with 3 mL of M9 minimal medium (47.7 mM Na_2HPO_4, 22.1 mM KH_2PO_4, 8.6 mM NaCl, 18.7 mM NH_4Cl, 2 mM $MgSO_4$, and 0.1 mM $CaCl_2$) supplemented with 0.4% glucose and 100 mg/l carbicillin, and 10 μM isopropyl- -D-thiogalactopyranoside (IPTG) was added to induce acyl-ACP TE expression. After 40 h of cultivation, cells were pelleted, and free fatty acids in the supernatant were extracted essentially following a previously described method (Voelker et al. 1994; Mayer et al. 2007). Briefly, 2 mL of culture supernatant was supplemented with 10 μg heptanoic acid (7:0), 10 μg undecanoic acid (11:0), and 20 μg heptadecanoic acid (17:0) (Sigma-Aldrich, St. Louis, MO, USA) as internal standards. The mixture was acidified with 20 μL of 1 M HCl, and 4 mL chloroform-methanol (1:1 vol/vol) was used to recover the fatty acids from the medium. After vortexing for 10 min and centrifuging at 1000 x *g* for 4 min, the lower chloroform phase was transferred to a new tube and evaporated under a stream of N_2 gas until the samples were concentrated to ~300 μL. Samples (1 μL) were analyzed on an Agilent Technologies (Santa Clara, CA, USA) 6890 Series GC system used with an Agilent 5973 mass selective detector equipped with an Agilent CP-Wax 58 FFAP CB column (25 m x 0.15 mm x 0.39 mm). The GC program followed an initial temperature of 70 ℃ for 2 min, ramped to 150 ℃ at 10 ℃/min and held for 3 min, ramped to 260 ℃ at 10 ℃/min, and held for 14 min. Final quantification analysis was performed with AMDIS software (National Institute of

Standards and Technology). Determination of C4 to C8, C10 to C12, and >C12 fatty acid concentrations was based on the fatty acid internal standards 7:0, 11:0, and 17:0, respectively. The total concentration of fatty acids produced by each acyl-ACP TE was obtained by subtracting the concentration of fatty acid produced by *E. coli* expressing a control plasmid (pUC57) lacking a TE from that produced by *E. coli* expressing a given acyl-ACP TE sequence from the same vector.

Identification of methylketone 2-tridecanone

Possible peaks for 2-tridecanone were observed when analyzing free fatty acids. To make further identification, 2-tridecanone standard purchased from Sigma-Aldrich was run on GC-MS with the same conditions as described above and retention time and MS spectrum of the peaks in each sample were compared to 2-tridecanone standard.

Statistical cluster analysis

To classify acyl-ACP TEs based on their *in vivo* activities, the fatty acid composition data obtained from the *in vivo* expression of all TE sequences studied herein were used to perform statistical clustering analysis. The distance matrix was calculated using Euclidean distances, and Ward's method (Ward 1963) was used to perform agglomerative hierarchical clustering. The *p*-values were calculated via multiscale bootstrap resampling with 1000 replicates (Suzuki et al. 2006).

Results

Two complementary approaches were taken to understand the breadth of substrate specificities exhibited by acyl-ACP TEs isolated from different taxa. In the first approach, we used phylogenetic analysis of all Family TE14 members of known or predicted function to strategically choose diverse TE sequences that were then expressed and functionally characterized. In the second approach, previously uncharacterized acyl-

ACP TEs were cloned from seeds of plants known to produce seed oils containing SCFAs and MCFAs.

Phylogenetic analysis and identification of acyl-ACP TEs

A total of 360 amino acid sequences belonging to Family TE14 (Cantu et al. 2010) were subjected to phylogenetic analysis and grouped into subfamilies. A subfamily is defined as having at least five sequences from different species, and it must pass the statistical tests described in the experimental procedures. Ten subfamilies met these criteria (Figure 2.1), accounting for 326 TE sequences; in addition 34 TE sequences could not be grouped into any of these subfamilies. All *z*-values were >3.4, ranging from 3.41 to 29.7, and mean distances between different subfamilies were larger than those within subfamilies.

Family TE14 contains acyl-ACP TEs that had previously been characterized from plants and classified into two types, FatA and FatB (Jones et al. 1995). Of the ten subfamilies identified in this study, Subfamilies A, B, and C are comprised of acyl-ACP TEs found in plants. All experimentally characterized sequences previously classified as FatB acyl-ACP TEs make up ~25% of Subfamily A , which contains 81 angiosperm-sourced sequences. The coconut and *C. viscosissima* acyl-ACP TEs identified in this study also belong to this subfamily. Subfamily B, which comprises 21 sequences primarily sourced from angiosperms as well as from the moss *Physcomitrella patens*, represents a potentially novel plant acyl-ACP TE subfamily with no previous experimental or phylogenetic characterization. Plant FatA acyl-ACP TEs, which act on long-chain acyl-ACP molecules, especially oleoyl-ACP (Jones et al. 1995), belong to the 32-member Subfamily C. As with Subfamily B, the six green algal sequences from *Chlamydomonas, Ostreococcus,* and *Micromonas* that comprise Subfamily D have not been experimentally characterized.

Unlike several plant acyl-ACP TEs, no bacterial acyl-ACP TEs had been functionally characterized. A total of 186 bacterial acyl-ACP TE sequences were classified into six subfamilies (Subfamily E–Subfamily J). All 17 acyl-ACP TE

sequences from gram-negative bacteria are in Subfamily E, which includes sequences from halophilic (*Salinibacter* and *Rhodothermus*), sulfate-reducing (*Desulfovibrio*, *Desulfohalobium*, and *Desulfonatronospira*), chemoorganotrophic (*Spirosoma*), metal-reducing (*Anaeromyxobacter*, *Geobacter*, and *Pelobacter*), and marine (*Microscilla*) bacteria. Subfamily F consists of 24 sequences, mainly from *Bacteroides* but also from other related bacteria. Protein Data Bank (PDB) structure 2ESS (Figure 2.2), obtained from a structural genomic effort, is part of this subfamily. Subfamily G and Subfamily H have 31 and 27 sequences, respectively, primarily from *Clostridium*. Subfamily I is comprised of eight sequences from six genera. Gram-positive lactic acid bacteria, almost completely from the genera *Lactobacillus*, *Enterococcus*, and *Streptococcus*, are part of Subfamily J. PDB:2OWN (Figure 2.2), the second bacterial acyl-ACP TE structure obtained from a structural genomic effort, appears in this family. Although the two known Family TE14 crystal structures (PDB:2ESS in Subfamily F and PDB:2OWN in Subfamily J) are from organisms in widely separated subfamilies, they are highly similar, as may be expected since they are members of the same enzyme family (Figure 2.2).

Some Family TE14 sequences are not grouped into any subfamily because their inclusion decreased *z*-values below acceptable limits. These include two plant and four moss sequences adjacent to Subfamilies A and C, and 28 bacterial sequences more closely related to Subfamilies E to I. No experimental work had previously been done on any of these sequences.

Upon generating the phylogenetic relationships among the 360 acyl-ACP TE sequences predicted or experimentally placed in Family TE14, 25 were chosen for experimental characterization. Of these, the cDNA for 24 was synthesized, while the cDNA of the *Elaeis guineensis* (oil palm) acyl-ACP TE was isolated from a phage cDNA library previously constructed from mRNA isolated from the developing fruit of Indonesian-sourced oil palm.

The selection of acyl-ACP TEs to characterize was based upon the primary structure-based phylogenetic relationships among the enzymes, along with knowledge of the fatty acid profile of the source organisms of these acyl-ACP TEs. Briefly, at least one

TE was characterized from each of the ten subfamilies except for Subfamily C, whose members appear to be specific for oleoyl-ACP substrates. For subfamilies that contain acyl-ACP TEs originating from organisms without any known fatty acid data, or from organisms where acyl-ACP TEs were not previously characterized, we chose to investigate acyl-ACP TE sequences that are evolutionarily distant from each other within each subfamily. For example, within Subfamily A there are two distinct and separate groupings of acyl-ACP TEs that are derived from the Poaceae family, for which there is no functional characterization (Table 2.1). One grouping contains one sorghum acyl-ACP TE sequence (GenBank:EER87824) and the other contains two (GenBank:EER88593 and GenBank:EES04698). To explore this structural divergence as an indicator of potential functional divergence in substrate specificities, one each of these Subfamily A sorghum acyl-ACP TEs (GenBank:EER87824 and GenBank:EER88593) and the two Subfamily B sorghum acyl-ACP TEs were expressed and functionally characterized.

Isolation and sequence analysis of acyl-ACP TEs from coconut and C. viscosissima

MCFAs are abundant in the oil produced in fruits of coconut (i.e. predominantly C12 and C14 and a small amount (0.2–1%) of C6 fatty acids (Kumar et al. 2004; Kumar et al. 2006; Kumar et al. 2009) and seeds of *C. viscosissima* (i.e. predominantly C8 and C10 fatty acids (Phippen et al. 2006)). Therefore, acyl-ACP TEs in the seeds of these species are predicted to be specific for medium-chain acyl-ACPs. Acyl-ACP TE sequences were isolated from coconut and *C. viscosissima* by a homologous cloning strategy. Using degenerate primers, which were designed from conserved regions of plant TE14 family enzymes, a 350-bp fragment in the middle of the mRNAs was amplified from cDNA generated from both developing coconut endosperm and *C. viscosissima* seeds. Sequencing of cloned PCR products identified three new acyl-ACP TE sequences each from coconut and *C. viscosissima*. The full-length cDNA sequences were obtained by RACE for three acyl-ACP TEs [CnFatB1 (JF338903), CnFatB2

(JF338904), and CnFatB3 (JF338905)] from coconut and three [CvFatB1 (JF338906), CvFatB2 (JF338907), and CvFatB3 (JF338908)] from *C. viscosissima*.

The predicted open reading frames of coconut and *C. viscosissima* acyl-ACP TE cDNAs were identified. They encode pre-proteins of 412 to 423 amino acids, with calculated molecular weights of 45.8 to 46.5 kDa and theoretical pIs of 6.4 to 8.8. Plant acyl-ACP TEs are nuclear- encoded, plastid-targeted proteins with an N-terminal plastid-targeting peptide extension (Voelker et al. 1992). For each of the cloned coconut and *C. viscosissima* acyl-ACPs TEs, the putative plastid-targeting peptide cleavage site was located on the N-terminal side of the conserved sequence LPDW (Figure 2.3), as proposed for many other plant acyl-ACP TEs (Dormann et al. 1995; Jones et al. 1995; Jha et al. 2006; Sanchez-Garcia et al. 2010; Moreno-Perez et al. 2011). These yield predicted mature proteins of 323 to 331 amino acid residues (Huynh et al. 2002), with calculated molecular weights of 36.6 to 37.5 kDa and theoretical pIs of 5.4 to 7.3. Alignment of the deduced amino acid sequences of coconut and *C. viscosissima* acyl-ACP TEs showed that, except for the plastid-targeting peptide sequences and very near the C-terminus, the sequences are colinear and share very high identity (63–86%) within a species (Figure 2.3). These sequences cluster within Subfamily A.

Determination of in vivo activities of acyl-ACP TEs

All isolated acyl-ACP TEs were expressed in *E. coli* strain K27. Secreted fatty acids were analyzed with GC-MS, and the total fatty acid yield in the medium was used to represent the *in vivo* activities of these enzymes on acyl-ACPs, though it remains possible that some of these enzymes might also cleave acyl-CoAs.

A total of 13 acyl-ACP TEs from Subfamily A were characterized, including single acyl-ACP TEs from *Cuphea palustris* (GenBank:AAC49179), *U. americana* (GenBank:AAB71731), and oil palm (*E. guineensis*, GenBank:AAD42220), two each from *Iris germanica* (GenBank:AAG43857 and GenBank:AAG43858) and *Sorghum bicolor* (GenBank:EER87824 and GenBank:EER88593), and three each from coconut and *C. viscosissima*. Total fatty acid concentrations produced by these acyl-ACP TEs are

listed in Table 2.1 and 2.2, and the resulting fatty acid compositions are shown in Figure 2.4. Acyl-ACP TEs from *C. palustris* and *U. americana*, which have previously been functionally characterized *in vitro* (Dehesh et al. 1996; Voelker et al. 1997), were studied as controls.

C. palustris acyl-ACP TE produced 97 mol% 8:0 and only 0.8 mol% 10:0 fatty acids (Figure 2.4A), while *U. americana* acyl-ACP TE made 44 mol% 8:0 and 23 mol% 10:0 fatty acids (Figure. 2.4B). *E. guineensis* acyl-ACP TE produced mainly 14:0 (47 mol%) and 16:1 (26 mol%) fatty acids (Figure. 2.4C). The acyl-ACP TEs from *I. germanica* and *S. bicolor* have similar substrate specificities, producing mainly 14:0 (30–46 mol%), 16:0 (11–23 mol%), and 16:1 (31–44 mol%) fatty acids (Figure. 2.4B). CnFatB1 (JF338903) and CnFatB2 (JF338904) made predominantly 14:0 (36–44 mol%) and 16:1 (31–44 mol%) fatty acids, whereas CnFatB3 (JF338905) made mainly 12:0 (34 mol%) and 14:1 (22 mol%) fatty acids (Figure. 2.4C). Finally, CvFatB1 (JF338906) produced mainly 8:0 (51 mol%) and 10:0 (25 mol%), and CvFatB2 (JF338907) made mainly 14:0 (46 mol%), 16:0 (25 mol%) and 16:1 (20 mol%) fatty acids (Figure. 2.4A). In contrast, CvFatB3 (JF338908) has narrower substrate specificity, producing predominantly 14:0 fatty acid (84 mol%).

Three acyl-ACP TEs from plant sources belonging to Subfamily B, including those from *P. patens* (GenBank:EDQ65090) and *S. bicolor* (GenBank:EER96252 and GenBank:EES11622), and one acyl-ACP TE from Subfamily D sourced from the alga *Micromonas pusilla* (GenBank:EEH52851), were similarly characterized. Total fatty acid production in *E. coli* expressing these acyl-ACP TEs varied from 9 to 380 nmol/mL (Table 2.1 and 2.2). These four acyl-ACP TEs showed similar substrate specificities, producing predominantly 14:0 (34–65 mol%) and 16:1 (23–37 mol%) fatty acids (Figure. 2.4D).

Eleven acyl-ACP TE sequences from Subfamilies E to J sourced from bacteria and three bacterial sequences that were not placed in any subfamily were characterized (Table 2.1, Table 2.2, and Figure. 2.5). Based on their substrate specificities, these acyl-ACP TEs were classified into two groups. One group produced primarily SCFAs and

MCFAs (>75 mol% 4:0 to 8:0 fatty acids). This group included acyl-ACP TEs from *Anaerococcus tetradius* (GenBank:EEI82564, no subfamily, 87% 8:0), *Clostridium perfringens* (GenBank:ABG82470, Subfamily G, 14% 6:0 and 70% 8:0), *Lactobacillus brevis* (GenBank:ABJ63754, Subfamily J, 7% 4:0, 14% 6:0, and 55% 8:0), and *Lactobacillus plantarum* (GenBank:CAD63310, Subfamily J, 11% 6:0 and 68% 8:0) (Figure. 2.5 and Table 2.2). The other group showed broad- and binary-range substrate specificities. The binary-range activities were centered on C8 and C12/C14 substrates (Figure. 2.5). Interestingly, many bacterial acyl-ACP TEs, such as those from *Desulfovibrio vulgaris* (GenBank:ACL08376, Subfamily E), *L. brevis* (GenBank:ABJ63754, Subfamily J), *L. plantarum* (GenBank:CAD63310, Subfamily J), and *Bdellovibrio bacteriovorus* (GenBank:CAE80300, no subfamily), are part of the pathway that produces noticeable amounts of the methylketone 2-tridecanone through enzymatic hydrolysis of 3-keto-tetradecanoyl-ACP followed by chemical decarboxylation (data not shown). *B. bacteriovorus* acyl-ACP TE produced the highest concentration of 2-tridecanone, 9.4 nmol/mL (Figure 2.6), which was 3 mol% of the fatty acids produced.

Clustering acyl-ACP TEs based on their catalytic functionality

To classify acyl-ACP TEs based on their substrate specificities, cluster analysis was performed on the fatty acid composition data as described in the Experimental Procedures. All acyl-ACP TEs characterized in this study clustered into three classes: 1) Class I contains acyl-ACP TEs that mainly act on C14 and C16 substrates; 2) Class II has acyl-ACP TEs that have broad substrate specificities, with major activities toward C8 and C14 substrates; and 3) Class III comprises acyl-ACP TEs that predominantly act on C8 substrate (Figure 2.7). Class I consists of thirteen plant acyl-ACP TEs from Subfamilies A, B, and D. Class II contains eleven acyl-ACP TEs, ten from bacteria in Subfamilies E, F, H, I, and J, and a non-grouped sequence, and only one from a plant (CnFatB3) in Subfamily A. Class III includes seven acyl-ACP TEs, of which three are from plants in Subfamily A and four are from bacteria in Subfamilies G and J and a non-

grouped sequence. Considering the previously characterized class of oleoyl-ACP TEs in Subfamily C, TE14 members may now be sorted into four classes based on their substrate specificities.

Discussion

The systematic functional characterization of bacterial acyl-ACP TEs demonstrates production of SCFAs

Over the past few decades, the number of acyl-ACP TE sequences in public databases has increased exponentially. The vast majority of these annotations are based solely on primary sequence homology; most have not been functionally characterized. The difficulty of purifying protein and preparing substrates precludes a large-scale *in vitro* characterization of acyl-ACP TEs. However, the well-known and widely used approach of analyzing fatty acid concentrations and distributions produced by heterologous TEs expressed in *E. coli* K27 provided an efficient and faster way to study the activities of a large number of diverse acyl-ACP TEs. Moreover, the integration of phylogeny and prior knowledge of the fatty acid profiles of the source organisms for these enzymes allowed us to rationally choose a representative subset of 31 acyl-ACP TEs to characterize. Significantly, this study represents the first experimental validation and functional characterization of bacterial acyl-ACP TEs, 14 of which were studied here.

Seven of these bacterial acyl-ACP TEs, those from *Bacteroides fragilis* (GenBank:CAH09236, Subfamily F), *B. thetaiotaomicron* (GenBank:AAO77182, Subfamily F), *Clostridium asparagiforme* (GenBank:EEG55387, Subfamily H), *Bryantella formatexigen* (GenBank:EET61113, Subfamily H), *L. brevis* (GenBank:ABJ63754, Subfamily J), *L. plantarum* (GenBank:CAD63310, Subfamily J), and *Streptococcus dysgalactiae* (GenBank:BAH81730, Subfamily J) produced significant amounts of 4:0 and 6:0 fatty acids when expressed in *E. coli*. This is the first report of acyl-ACP TEs that have these catalytic activities. Although these enzymes did

not appear to show high activities against C4-ACP and C6-ACP, they provide a good starting point for protein engineering. Both of these SCFAs could then be potential candidates for platform biochemicals for a biorenewable chemical industry (Nikolau et al. 2008).

Acyl-ACP TEs from MCFA-producing plant tissues make MCFAs

Plant tissues known to produce MCFAs were shown in this study to contain at least one acyl-ACP TE specific for medium chain acyl-ACPs. It appears that CnFatB1 (producing primarily 8:0, 14:0, and 16:0 fatty acids), CnFatB2 (making mainly 14:0, 16:0, and 16:1 fatty acids) and CnFatB3 (producing mostly 8:0, 12:0, 14:0, and 14:1 fatty acids) might work together to determine the fatty acid composition of coconut oil, which contains primarily 12:0 (43–50%) and 14:0 (16–22%) and small amounts of 6:0, 8:0, and 10:0 fatty acids (Kumar et al. 2004; Kumar et al. 2006; Kumar et al. 2009). However, we cannot rule out the possibility that other acyl-ACP TEs are also expressed in coconut endosperm, and that they may be involved in determining the fatty acid composition of the oil.

CvFatB1 and CvFatB3 produced MCFAs in *E. coli*, and CvFatB1 shows substrate specificity consistent with the fatty acid constituents present in the seed oil. The relative distributions of 8:0 and 10:0 fatty acids differ; CvFatB1 produced twice as much 8:0 compared to 10:0 fatty acid, whereas there is ~fourfold more 10:0 fatty acid within *C. viscosissima* seed oil (Phippen et al. 2006). Differences in *in vivo* substrate activities in *E. coli* K27 compared to *in vitro* enzymatic assays, or in the fatty acid composition of the organism from which the acyl-ACP TE was sourced, have been noted previously (Dehesh et al. 1996; Dehesh et al. 1996; Leonard et al. 1997; Voelker et al. 1997). This phenomenon reflects the complexity of the fatty acid biosynthesis pathway within the plant. For example, multiple acyl-ACP TEs within an organism may contribute to fatty acid composition. Alternatively, the fatty acid profile of an organism may be determined by the kinetics of the entire fatty acid biosynthesis pathway, as has been previously proposed (Davies 1993; Voelker et al. 1994; Leonard et al. 1997),

including the contribution made by the species-specific interactions between the acyl-ACP TE and the ACP molecule that carries the acyl-substrate for the acyl-ACP TE. Regardless, this study identifies specific medium-chain substrates on which the TE can act, which is especially important for engineering the fatty acid biosynthesis pathway.

Acyl-ACP TEs can intercept both saturated and unsaturated intermediates of Type II fatty acid synthase of E. coli

Several plant acyl-ACP TEs (e.g. CnFatB3) produced significant amounts of unsaturated fatty acids (UFAs) when expressed in *E. coli*. These include 10:1, 12:1, 14:1, and 16:1 fatty acids (Figure 2.4), which do not usually accumulate in *E. coli* or in the original host plant tissues from which the acyl-ACP TE was isolated. A similar finding has been reported for a *U. californica* acyl-ACP TE expressed in *E. coli* K27 (Voelker et al. 1994). Although the double bond position within these fatty acids was not determined in this study, double bonds in UFAs produced in *E. coli* K27 expressing a *Cinnamonum camphorum* acyl-ACP TE were all in *cis* conformation and at the – 7 position (Feng et al. 2009). *E. coli* has a different UFA biosynthesis pathway than plants. Bacteria, such as *E. coli*, utilize an anaerobic system different in which the double bond is retained in the acyl chain as it is being assembled. Plants instead use aerobic acyl-ACP desaturase to introduce double bonds into the acyl chain once it is preformed. Specifically, the FabA gene in *E. coli*, encoding 3-hydroxydecanoyl-ACP dehydratase/isomerase, is a bifunctional enzyme that introduces a double bond at C10 and regulates the branch point of the saturated and unsaturated pathways (Magnuson et al. 1993). FabB encodes a 3-ketoacyl-ACP synthase that catalyzes the elongation of *cis*-3-decenoyl-ACP produced by FabA (Magnuson et al. 1993). Because 10:1-ACP, 12:1-ACP, 14:1-ACP, and 16:1-ACP are intermediates of the UFA biosynthesis pathway in *E. coli*, the UFAs produced by acyl-ACP TEs are most likely derived from those intermediates and thus are in the *cis* conformation and unsaturated at the – 7 position. The accumulation of both UFAs and saturated fatty acids observed in this study is consistent with previous conclusion

that the heterologously expressed acyl-ACP TEs can intercept both saturated and unsaturated intermediates of fatty acid biosynthesis of *E. coli* [44].

Subtle changes in primary sequences may be sufficient to change the substrate specificity of acyl-ACP TEs

The relationship between the structures of acyl-ACP TEs and their functionalities (i.e. their substrate specificities) is poorly understood. To begin to address this question, the 31 acyl-ACP TEs that were functionally characterized herein were clustered using the substrate specificity data obtained from their *in vivo* activities (Figures. 2.4 and 2.5 and Table 2.2). Comparison between the specificity-based classification and the sequence-based phylogenetic tree (Figure. 2.1) indicates that the two classifications are not necessarily consistent with each other. Three phenomena were observed in this study. First, diverged sequences (variants in primary structure) from the same species do not necessarily differ in function. For example, *S. bicolor* expresses at least three acyl-ACP TEs in Subfamily A and two in Subfamily B, all of which share very similar substrate specificity as measured by the fatty acids produced when expressed in *E. coli* (Figure 2.4D). One possible explanation for the persistence of this number of acyl-ACP TEs with similar function within a species genome may be due to divergence in spatial or temporal expression of their acyl-ACP TEs. Second, similar sequences may have different substrate specificities, e.g., three acyl-ACP TEs from *C. viscosissima* have different substrate specificities although their mature protein sequences share more than 70% primary sequence identity, and they all classified within Subfamily A. Third, sequences that belong to different subfamilies because they share low sequence identity can have very similar substrate specificities. For example, CnFatB2 (Subfamily A) and *S. bicolor* (GenBank:EER96252, Subfamily B) acyl-ACP TEs are members of different subfamilies and share only 40% sequence identity, and yet they have very similar substrate specificities. Therefore, it is not reasonable to infer the substrate specificity of one acyl-ACP TE based on its sequence -based classification within the same subfamily. It is conceivable, therefore, that the change of substrate specificity is most likely caused

by changes of only a few amino acid residues, and that many different combinations of residue changes could result in changed specificities (Jones et al. 1995). Identifying the amino acids that determine substrate specificity is critical for engineering novel acyl-ACP TEs, but this is limited by the lack of tertiary structural information of acyl-ACP TEs from different subfamilies. A comparison of the two PDB structures known for bacterial acyl-ACP TEs, from *B. thetaiotamicron* (PDB:2ESS, GenBank:AAO77182, Subfamily F) and *L. plantarum* (PDB:2OWN, GenBank:CAD63310, Subfamily J), is instructive. Although they share only 18% sequence identity, these two proteins share a common HotDog tertiary structure, being co-aligned with an RMSD of 2.59 Å (Figure 2.2). However, *B. thetaiotamicron* acyl-ACP TE has broad substrate specificity, while *L. plantarum* acyl-ACP TE is specific for C6 and C8 acyl-ACP substrates. Thus, future work can focus on identifying and validating the role of specific residues in determining acyl-ACP TE substrate specificity.

Unexpected activity reveals diversity of acyl-ACP TEs

Interestingly, in the *E. coli* heterologous expression system used here, six bacterial-sourced acyl-ACP TEs and three plant-sourced acyl-ACP TEs produced noticeable amounts (>1 nmol/mL) of methylketones, largely 2-tridecanone. The acyl-ACP TE from *B. bacteriovorus* (GenBank:CAE80300) produced the highest concentration of 2-tridecanone (9.4 nmol/mL).

Methylketones such as 2-tridecanone occur in the wild tomato species *Solanum habrochaites* subsp. *Glabratum* (Antonious 2001), and their biosynthesis is catalyzed by two sequentially-acting methylketone synthases, MKS1 and MKS2. MKS2 is a TE that catalyzes the hydrolysis of the 3-ketoacyl-ACP intermediate in fatty acid biosynthesis, and MKS1 catalyzes the decarboxylation of the released 3-keto acid to produce a methylketone (Ben-Israel et al. 2009; Yu et al. 2010). Heterologous expression of MKS2 in *E. coli* yields many methylketones, including 2-tridecanone (Yu et al. 2010). However, MKS2 is not included in Family TE14, but instead it belongs to Family TE9 (Cantu et al. 2010). Although some Family TE14 members share very low if any significant sequence

similarity (i.e., <15% identity) to MKS2, the current study indicates that at least nine acyl-ACP TEs (e.g. *B. bacteriovorus*, GenBank:CAE80300) can catalyze the same reaction as MKS2 (i.e, hydrolysis of the thioester bond of 3-ketoacyl-ACP), and that the resulting product (3-keto acid) is further chemically or enzymatically decarboxylated to generate the methylketone.

Conclusions

This study has revealed that acyl-ACP TEs isolated from different taxa have considerable functional diversity relative to their substrate specificity. Prior characterizations of plant acyl-ACP TEs have focused on the substrate specificity relative to acyl chain lengths, to identify such enzymes for bioengineering a source of lauric acid for use by the detergent industry. The present study has revealed that bacterial orthologs provide access to additional functional diversity, both relative to acyl chain length specificity (e.g., shorter acyl chains, as short as four carbon atoms), as well as acyl chains that contain additional chemical functionalities (e.g., unsaturated acyl chains and acyl chains containing carbonyl groups). This additional functional diversity in acyl-ACP TEs can potentially be used to diversify the fatty acid biosynthesis pathway to produce biorenewable chemicals (Nikolau et al. 2008).

Acknowledgments

This work was supported by the U.S. National Science Foundation through its Engineering Research Center Program (Award No. EEC-0813570), leading to the Center for Biorenewable Chemicals (CBiRC), headquartered at Iowa State University and including Rice University, the University of California, Irvine, the University of New Mexico, the University of Virginia, and the University of Wisconsin-Madison. The authors thank Professor Derrick Rollins (Iowa State University) for supplying the equation to establish the statistical justification for separating the subfamilies, the USDA-ARS-SHRS National Germplasm Repository for providing coconut fruits, Laura Marek and Irvin Larsen at the North Central Regional Plant Introduction Station for

providing cuphea seeds, field space, and helpful expertise, Sumira Stein for assistance with experiments, and M. Ann D.N. Perera of the W.M. Keck Metabolomics Research Laboratory at Iowa State University for assistance with fatty acid analysis. We also thank Asmini Budiani (Bogor Agricultural University, Bogor, Indonesia) for preparing the phage-based oil-palm cDNA library during a four-month visit to the Nikolau laboratory.

Authors' contributions

FJ and DCC contributed equally to the work. This body of work represents a collaboration between the Reilly and Nikolau laboratories. Graduate student DCC and undergraduate JPC conducted the computational and phylogenetic research in the Reilly laboratory. In the Nikolau laboratory graduate student FJ and Ames High School (Ames, IA) student JT performed the molecular and biochemical research with the aid and supervision of research scientist MDY-N. FJ, DCC, BJN, MDY-N, and PJR wrote the manuscript.

References

Antonious, G. F. (2001). Production and quantification of methyl ketones in wild tomato accessions. J Environ Sci Health B 36(6): 835-848.

Ben-Israel, I., Yu, G., Austin, M. B., Bhuiyan, N., Auldridge, M., Nguyen, T., Schauvinhold, I., Noel, J. P., Pichersky, E. and Fridman, E. (2009). Multiple biochemical and morphological factors underlie the production of methylketones in tomato trichomes. Plant Physiology 151(4): 1952-1964.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. and Sayers, E. W. (2011). GenBank. Nucleic Acids Res 39(suppl 1): D32-D37.

Cantu, D. C., Chen, Y., Lemons, M. L. and Reilly, P. J. (2011). ThYme: a database for thioester-active enzymes. Nucleic Acids Res 39(Database issue): D342-346.

Cantu, D. C., Chen, Y. F. and Reilly, P. J. (2010). Thioesterases: A new perspective based on their primary and tertiary structures. Protein Science 19(7): 1281-1295.

Consortium, U. (2010). The Universal Protein Resource (UniProt) in 2010. Nucleic Acids Research 38: D142-D148.

Davies, H. M. (1993). Medium-Chain Acyl-Acp Hydrolysis Activities of Developing Oilseeds. Phytochemistry 33(6): 1353-1356.

Dehesh, K., Edwards, P., Hayes, T., Cranmer, A. M. and Fillatti, J. (1996). Two novel thioesterases are key determinants of the bimodal distribution of acyl chain length of Cuphea palustris seed oil. Plant Physiology 110(1): 203-210.

Dehesh, K., Jones, A., Knutzon, D. S. and Voelker, T. A. (1996). Production of high levels of 8:0 and 10:0 fatty acids in transgenic canola by overexpression of Ch FatB2, a thioesterase cDNA from Cuphea hookeriana. Plant Journal 9(2): 167-172.

Desper, R. and Gascuel, O. (2002). Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. Journal of Computational Biology 9(5): 687-705.

Dormann, P., Voelker, T. A. and Ohlrogge, J. B. (1995). Cloning and Expression in Escherichia-Coli of a Novel Thioesterase from Arabidopsis-Thaliana Specific for Long-Chain Acyl-Acyl Carrier Proteins. Archives of Biochemistry and Biophysics 316(1): 612-618.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32(5): 1792-1797.

Feng, Y. J. and Cronan, J. E. (2009). Escherichia coli Unsaturated Fatty Acid Synthesis COMPLEX TRANSCRIPTION OF THE fabA GENE AND IN VIVO IDENTIFICATION OF THE ESSENTIAL REACTION CATALYZED BY FabB. Journal of Biological Chemistry 284(43): 29526-29535.

Hawkins, D. J. and Kridl, J. C. (1998). Characterization of acyl-ACP thioesterases of mangosteen (Garcinia mangostana) seed and high levels of stearate production in transgenic canola. Plant Journal 13(6): 743-752.

Huynh, T. T., Pirtle, R. M. and Chapman, K. D. (2002). Expression of a Gossypium hirsutum cDNA encoding a FatB palmitoyl-acyl carrier protein thioesterase in Escherichia coli. Plant Physiology and Biochemistry 40(1): 1-9.

Jha, J. K., Maiti, M. K., Bhattacharjee, A., Basu, A., Sen, P. C. and Sen, S. K. (2006). Cloning and functional expression of an acyl-ACP thioesterase FatB type from Diploknema (Madhuca) butyracea seeds in Escherichia coli. Plant Physiology and Biochemistry 44(11-12): 645-655.

Jones, A., Davies, H. M. and Voelker, T. A. (1995). Palmitoyl-Acyl Carrier Protein (Acp) Thioesterase and the Evolutionary Origin of Plant Acyl-Acp Thioesterases. Plant Cell 7(3): 359-371.

Jones, D. T., Taylor, W. R. and Thornton, J. M. (1992). The Rapid Generation of Mutation Data Matrices from Protein Sequences. Computer Applications in the Biosciences 8(3): 275-282.

Klein, K., Steinberg, R., Fiethen, B. and Overath, P. (1971). Fatty acid degradation in Escherichia coli. An inducible system for the uptake of fatty acids and further characterization of old mutants. Eur J Biochem 19: 442-450.

Kumar, S. N. and Balakrishna, A. (2009). Seasonal Variations in Fatty Acid Composition of Oil in Developing Coconut. Journal of Food Quality 32(2): 158-176.

Kumar, S. N., Balakrishnan, A. and Rajagopal, V. (2006). Fatty acids in coconut oil. Indian Coconut Journal 37(5): 4-14.

Kumar, S. N., Champakam, B. and Rajagopal, V. (2004). Variability in coconut cultivars for lipid and fatty acid composition of oil. Tropical Agriculture 81(1): 34-40.

Leonard, J. M., Slabaugh, M. B. and Knapp, S. J. (1997). Cuphea wrightii thioesterases have unexpected broad specificities on saturated fatty acids. Plant Molecular Biology 34(4): 669-679.

Magnuson, K., Jackowski, S., Rock, C. O. and Cronan, J. E. (1993). Regulation of Fatty-Acid Biosynthesis in Escherichia-Coli. Microbiological Reviews 57(3): 522-542.

Mayer, K. M. and Shanklin, J. (2007). Identification of amino acid residues involved in substrate specificity of plant acyl-ACP thioesterases using a bioinformatics-guided approach. Bmc Plant Biology 7.

Mertz, B., Kuczenski, R. S., Larsen, R. T., Hill, A. D. and Reilly, P. J. (2005). Phylogenetic analysis of family 6 glycoside hydrolases. Biopolymers 79(4): 197-206.

Moreno-Perez, S., Sanchez-Garcia, A., Salas, J. J., Garces, R. and Martinez-Force, E. (2011). Acyl-ACP thioesterases from macadamia (Macadamia tetraphylla) nuts: Cloning, characterization and their impact on oil composition. Plant Physiology and Biochemistry 49: 82-87.

Nikolau, B. J., Perera, M. A., Brachova, L. and Shanks, B. (2008). Platform biochemicals for a biorenewable chemical industry. Plant Journal 54(4): 536-545.

Overath, P., Pauli, G. and Schairer, H. U. (1969). Fatty acid degradation in Escherichia coli. An inducible acyl-CoA synthetase, the mapping of old-mutations, and the isolation of regulatory mutants. Eur J Biochem 7(4): 559-574.

Phippen, W. B., Isbell, T. A. and Phippen, M. E. (2006). Total seed oil and fatty acid methyl ester contents of Cuphea accessions. Industrial Crops and Products 24(1): 52-59.

Pollard, M. R., Anderson, L., Fan, C., Hawkins, D. J. and Davies, H. M. (1991). A Specific Acyl-Acp Thioesterase Implicated in Medium-Chain Fatty-Acid Production in Immature Cotyledons of Umbellularia-Californica. Archives of Biochemistry and Biophysics 284(2): 306-312.

Sanchez-Garcia, A., Moreno-Perez, A. J., Muro-Pastor, A. M., Salas, J. J., Garces, R. and Martinez-Force, E. (2010). Acyl-ACP thioesterases from castor (Ricinus communis L.): An enzymatic system appropriate for high rates of oil synthesis and accumulation. Phytochemistry 71(8-9): 860-869.

Serrano-Vega, M. J., Garces, R. and Martinez-Force, E. (2005). Cloning, characterization and structural model of a FatA-type thioesterase from sunflower seeds (Helianthus annuus L.). Planta 221(6): 868-880.

Suzuki, R. and Shimodaira, H. (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics 22(12): 1540-1542.

Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007). MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24(8): 1596-1599.

Voelker, T. A. and Davies, H. M. (1994). Alteration of the Specificity and Regulation of Fatty-Acid Synthesis of Escherichia-Coli by Expression of a Plant Medium-Chain Acyl-Acyl Carrier Protein Thioesterase. Journal of Bacteriology 176(23): 7320-7327.

Voelker, T. A., Jones, A., Cranmer, A. M., Davies, H. M. and Knutzon, D. S. (1997). Broad-range and binary-range acyl-acyl-carrier-protein thioesterases suggest an alternative mechanism for medium-chain production in seeds. Plant Physiology 114(2): 669-677.

Voelker, T. A., Worrell, A. C., Anderson, L., Bleibaum, J., Fan, C., Hawkins, D. J., Radke, S. E. and Davies, H. M. (1992). Fatty-Acid Biosynthesis Redirected to Medium Chains in Transgenic Oilseed Plants. Science 257(5066): 72-74.

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association 58(301): 236-&.

Yu, G., Nguyen, T. T. H., Guo, Y., Schauvinhold, I., Auldridge, M. E., Bhuiyan, N., Ben-Israel, I., Iijima, Y., Fridman, E., Noel, J. P. and Pichersky, E. (2010). Enzymatic Functions of Wild Tomato Methylketone Synthases 1 and 2. Plant Physiology 154(1): 67-77.

Yuan, L., Voelker, T. A. and Hawkins, D. J. (1995). Modification of the Substrate-Specificity of an Acyl-Acyl Carrier Protein Thioesterase by Protein Engineering. Proceedings of the National Academy of Sciences of the United States of America 92(23): 10639-10643.

Tables

Table 2.1 Total fatty acid production of synthesized and cloned acyl-ACP TEs

Kingdom	Subfamily	ACC No./Name	Organism	Rationale for synthesis[a]	Total FA[b] (nmol/mL)
Planta	A	AAC49179[c, d]	*Cuphea palustris*	A (Bimodal specificity for C8 and C10 substrates)	708 ± 45
		AAB71731	*Ulmus americana*	A (Broad specificity; highest activity on C10 and C16)	1098 ± 62
		AAG43857	*Iris germanica*	B	261 ± 20
		AAG43858	*Iris germanica*	B	14.8 ± 4.6
		EER87824	*Sorghum bicolor*	B (Member of a Subfamily A Poeceae TE cluster)	126 ± 13
		EER88593	*Sorghum bicolor*	B (Member of a Subfamily A Poeceae TE cluster)	90.7 ± 8.0
		CnFatB1	*Cocos nucifera*	C	130 ± 12
		CnFatB2	*Cocos nucifera*	C	572 ± 32
		CnFatB3	*Cocos nucifera*	C	200 ± 11
		CvFatB1	*Cuphea viscosissima*	C	79.2 ± 9.7
		CvFatB2	*Cuphea viscosissima*	C	249 ± 9
		CvFatB3	*Cuphea viscosissima*	C	18.9 ± 2.1
		AAD42220	*Elaeis guineensis*	C	36.7 ± 3.8
	B	EDQ65090	*Physcomitrella patens*	B (Member of novel plant subfamily)	380 ± 29

Table 2.1 continued

		EER96252	*Sorghum bicolor*	B (Member of novel plant subfamily)	175 ± 11
		EES11622	*Sorghum bicolor*	B (Member of novel plant subfamily)	9.43 ± 2.03
	D	EEH52851	*Micromonas pusilla*	B	16.3 ± 1.6
Bacteria	E	ACL08376	*Desulfovibrio vulgaris*	D (Medium-chain linear, branched, and hydroxy fatty acids)	330 ± 9
	F	CAH09236	*Bacteroides fragilis*	D (Hydroxy fatty acids)	215 ± 6
		ABR43801	*Parabacteroides distasonis*	D (Branched and branched hydroxy fatty acids)	70.3 ± 4.4
		AAO77182[e]	*Bacteroides thetaiotaomicron*	D (Anteiso-branched and hydroxy fatty acids)	60.4 ± 2.9
	G	ABG82470	*Clostridium perfringens*	D (Medium-chain fatty acids)	72.0 ± 9.5
	H	EEG55387	*Clostridium asparagiforme*	B	25.9 ± 4.2
		EET61113	*Bryantella formatexigens*	B	381 ± 3
	I	EDV77528	*Geobacillus sp.*	D (Iso-branched fatty acids)	64.9 ± 12.0
	J	BAH81730	*Streptococcus dysgalactiae*	D (Medium-chain and cyclic propane ring fatty acids)	623 ± 14
		ABJ63754	*Lactobacillus brevis*	D (Medium-chain and cyclic propane ring fatty acids)	710 ± 10
		CAD63310[e]	*Lactobacillus plantarum*	D (Medium-chain 3'-hydroxy fatty acids)	436 ± 10

Table 2.1 continued

	Non-grouped	EEI82564	*Anaerococcus tetradius*	D (Organism produces butyric acid)	1381 ± 146
		CAE80300	*Bdellovibrio bacteriovorus*	D (Straight-chain odd-numbered fatty acids)	333 ± 18
		ABN54268	*Clostridium thermocellum*	D (Branched-chain fatty acids)	97.7 ± 3.2

[a] A: Functionally characterized TEs; B: TE does not group near characterized TEs and/or no organism lipid profile information is available; C: TEs cloned from organisms known to produce MCFAs; D: Organism's lipid profile used and predominant fatty acid constituents identified in the organism are listed in parentheses.

[b] The data are represented as mean ± standard error (n = 4).

[c] All but the three *C. nucifera* sequences were codon-optimized for expression in *E. coli*.

[d] Transit peptides were removed from all plant sequences.

[e] Acyl-ACP TEs with known crystal structures.

TEs were expressed in *E. coli* K27, and free fatty acids (FAs) that accumulated in the medium were analyzed by GC-MS.

Table 2.2. Molar percentages and total concentrations of fatty acids produced by different TEs. a, TE genes were codon-optimized for the expression in *E. coli*; b, transit peptides of plant TEs were removed; c, TEs with known crystal structures; d, mean ± standard error (n=4)

Kingdom	Subfamily	ACC NO./ Name	Organism	Total FA[d] (nmol/ml)	Percentage of individual FA (mol %)										
					4:0	6:0	8:0	10:0	10:1	12:0	12:1	14:0	14:1	16:0	16:1
Planta	A	AAC49179[a,b]	*Cuphea palustris*	708±45	nt	0.24±0.05	97.5±0.2	0.85±0.08	0.11±0.02	0.64±0.06	0.17±0.01	0.15±0.08	0.20±0.04	nt	0.11±0.06
		AAB71731[a,b]	*Ulmus americana*	1098±62	0.11±0.03	0.40±0.05	44.2±3.1	23.4±1.4	3.22±0.44	3.72±0.37	7.87±0.94	9.80±0.77	1.53±0.18	1.39±0.14	4.40±0.68
		AAG43857[a,b]	*Iris germanica*	261±20	nt	0.01±0.09	3.29±0.43	0.53±0.05	0.07±0.01	0.74±0.10	0.95±0.08	29.8±3.0	0.31±0.10	20.1±2.9	44.2±3.9
		AAG43858[a,b]	*Iris germanica*	14.8±4.6	nt	nt	8.33±4.17	nt	nt	0.48±1.37	nt	32.2±13.3	nt	23.0±12.6	35.9±16.9
		EER87824[a,b]	*Sorghum bicolor*	126±13	nt	0.22±0.20	4.78±0.67	0.24±0.04	nt	1.58±0.32	0.35±0.06	45.6±5.2	nt	12.7±3.1	34.6±5.6
		EER88593[a,b]	*Sorghum bicolor*	90.7±8.0	nt	0.11±0.23	5.71±1.06	0.66±0.09	nt	3.16±0.48	0.59±0.26	45.0±4.6	3.28±0.90	10.5±2.2	31.0±3.8
		CnFatB1[b]	*Cocos nucifera*	130±12	0.08±0.07	0.39±0.22	14.4±2.9	1.12±0.25	0.13±0.04	1.35±0.34	1.05±0.12	44.0±4.5	0.36±0.27	6.04±1.37	31.1±4.6
		CnFatB2[b]	*Cocos nucifera*	572±32	0.04±0.02	0.09±0.04	1.69±0.21	0.05±0.02	0.01±0.00	0.95±0.15	0.34±0.05	35.7±3.0	0.73±0.11	16.0±1.7	44.4±2.7
		CnFatB3[b]	*Cocos nucifera*	200±11	nt	0.29±0.11	11.1±0.8	1.21±0.08	0.01±0.00	34.2±2.1	6.06±0.64	13.9±2.2	22.7±2.6	2.07±1.06	8.46±3.16
		CvFatB1[a,b]	*Cuphea viscosissima*	79.2±9.7	0.11±0.25	1.45±0.40	51.7±6.0	25.7±4.9	0.65±0.22	7.22±1.64	5.64±1.48	6.65±2.20	0.88±0.57	nt	nt
		CvFatB2[a,b]	*Cuphea viscosissima*	249±9	0.22±0.11	0.83±0.33	4.14±1.09	0.54±0.17	nt	1.00±0.31	0.93±0.39	46.7±1.8	0.71±0.51	25.5±1.5	19.4±2.3
		CvFatB3[a,b]	*Cuphea viscosissima*	18.9±2.1	nt	0.10±1.37	6.71±4.62	5.39±2.93	nt	2.23±1.87	1.67±1.49	83.9±5.6	nt	nt	nt
		AAD42220[a,b]	*Elaeis guineensis*	36.7±3.8	nt	3.13±1.14	14.3±3.8	1.93±0.80	nt	2.34±1.18	2.29±1.40	47.3±5.1	2.85±3.38	nt	26.0±4.9
	B	EDQ65090[a,b]	*Physcomitrella patens*	380±29	0.01±0.05	0.19±0.06	8.76±1.08	0.39±0.08	0.06±0.01	0.44±0.08	0.18±0.03	42.1±3.7	0.16±0.07	15.5±2.5	32.2±4.0
		EER96252[a,b]	*Sorghum bicolor*	175±11	nt	nt	5.80±0.88	1.79±0.81	0.07±0.04	1.22±0.33	1.05±0.37	33.7±2.9	0.93±0.19	18.2±2.5	37.3±3.2
		EES11622[a,b]	*Sorghum bicolor*	9.43±2.03	nt	0.23±3.60	5.32±2.58	nt	nt	nt	nt	50.6±10.8	nt	15.2±8.3	28.6±9.2
	D	EEH52851[a,b]	*Micromonas pusilla*	16.3±1.6	nt	nt	3.61±2.16	nt	nt	0.14±1.32	0.62±0.06	65.1±4.4	7.95±3.91	nt	22.6±2.5
Bacteria	E	ACL08376[a]	*Desulfovibrio vulgaris*	330±9	0.25±0.06	1.56±0.20	28.8±1.2	3.47±0.18	0.41±0.03	7.88±0.28	24.2±1.4	5.95±0.37	23.6±1.1	1.19±0.35	2.6±0.3
	F	CAH09236[a]	*Bacteroides fragilis*	215±6	13.1±0.8	2.72±0.23	20.2±1.5	2.66±0.22	0.32±0.07	3.64±0.38	19.2±1.4	5.05±0.30	25.4±1.2	2.23±0.34	5.44±0.40
		ABR43801[a]	*Parabacteroides distasonis*	70.3±4.4	0.40±0.25	1.29±0.74	18.0±4.5	6.28±0.39	nt	16.3±1.1	9.32±0.79	21.3±2.0	27.1±2.1	nt	nt
		AAO77182[a,c]	*Bacteroides thetaiotaomicron*	60.4±2.9	27.4±3.2	1.64±0.37	13.4±0.8	2.09±0.15	nt	4.59±0.68	16.7±0.9	6.06±1.08	25.6±1.4	nt	2.59±0.25
	G	ABG82470[a]	*Clostridium perfringens*	72.0±9.5	3.02±0.67	14.0±2.5	70.3±4.4	3.03±0.51	0.10±0.03	nt	1.05±0.19	nt	8.45±2.18	nt	nt
	H	EEG55387[a]	*Clostridium asparagiform*	25.9±4.2	1.92±0.69	4.45±1.52	26.0±5.7	5.51±1.41	nt	6.69±2.15	1.61±0.82	35.0±9.2	17.5±5.8	nt	1.27±0.87
		EET61113[a]	*Bryantella formatexigens*	381±3	15.0±0.3	20.4±0.2	31.8±0.3	5.08±0.13	0.70±0.02	4.30±0.15	8.88±0.46	1.85±0.19	10.5±0.2	0.40±0.24	1.17±0.11
	I	EDV77528[a]	*Geobacillus sp.*	64.9±12.0	2.35±1.70	1.09±0.57	8.58±4.44	2.42±1.08	2.57±1.43	6.76±3.13	30.8±10.6	11.0±3.4	31.7±8.2	0.30±1.11	2.37±1.22
	J	BAH81730[a]	*Streptococcus dysgalactiae*	624±14	3.92±0.18	13.2±0.6	29.9±1.1	5.02±0.18	0.80±0.02	5.73±0.33	13.5±0.7	4.44±0.25	20.0±1.1	0.25±0.14	3.30±0.14
		ABJ63754[a]	*Lactobacillus brevis*	710±10	7.05±0.29	13.7±0.3	55.5±0.7	2.58±0.06	0.73±0.01	3.75±0.07	7.94±0.19	1.85±0.14	6.27±0.19	nt	0.68±0.14
		CAD63310[a,c]	*Lactobacillus plantarum*	436±10	3.09±0.28	11.0±0.4	68.0±0.8	1.24±0.06	0.08±0.01	2.81±0.18	4.63±0.21	1.87±0.11	6.87±0.42	nt	0.45±0.06
	Non-grouped	EEI82564[a]	*Anaerococcus tetradius*	1381±146	0.53±0.07	1.35±0.19	86.7±1.5	2.18±0.30	0.46±0.08	1.13±0.18	2.79±0.48	1.18±0.23	3.01±0.54	0.05±0.09	0.66±0.11
		CAE80300[a]	*Bdellovibrio bacteriovorus*	333±18	0.10±0.09	0.86±0.14	36.9±3.1	3.26±0.45	0.44±0.07	6.65±0.66	7.56±0.64	8.20±0.73	27.8±2.1	1.60±0.22	6.57±0.49
		ABN54268[a]	*Clostridium thermocellum*	97.7±3.2	0.59±0.14	0.75±0.24	8.36±0.37	4.50±0.16	0.27±0.01	2.66±0.23	7.94±0.38	9.84±0.77	59.5±1.4	0.83±0.67	4.74±0.42

Figures

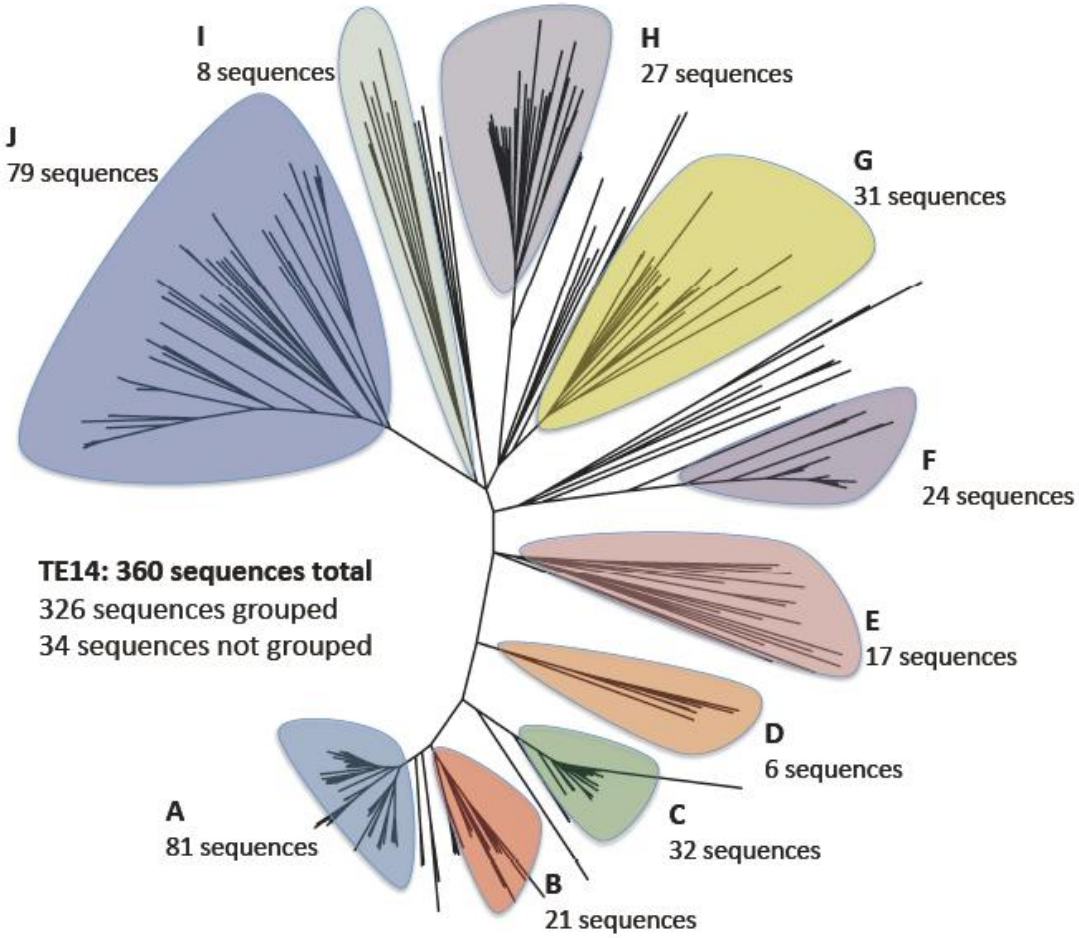


Figure 2.1 Unrooted phylogenetic tree of acyl-ACP TEs showing Subfamilies A to J. Those branches falling outside the shaded areas are non-grouped and therefore are not part of any subfamily.

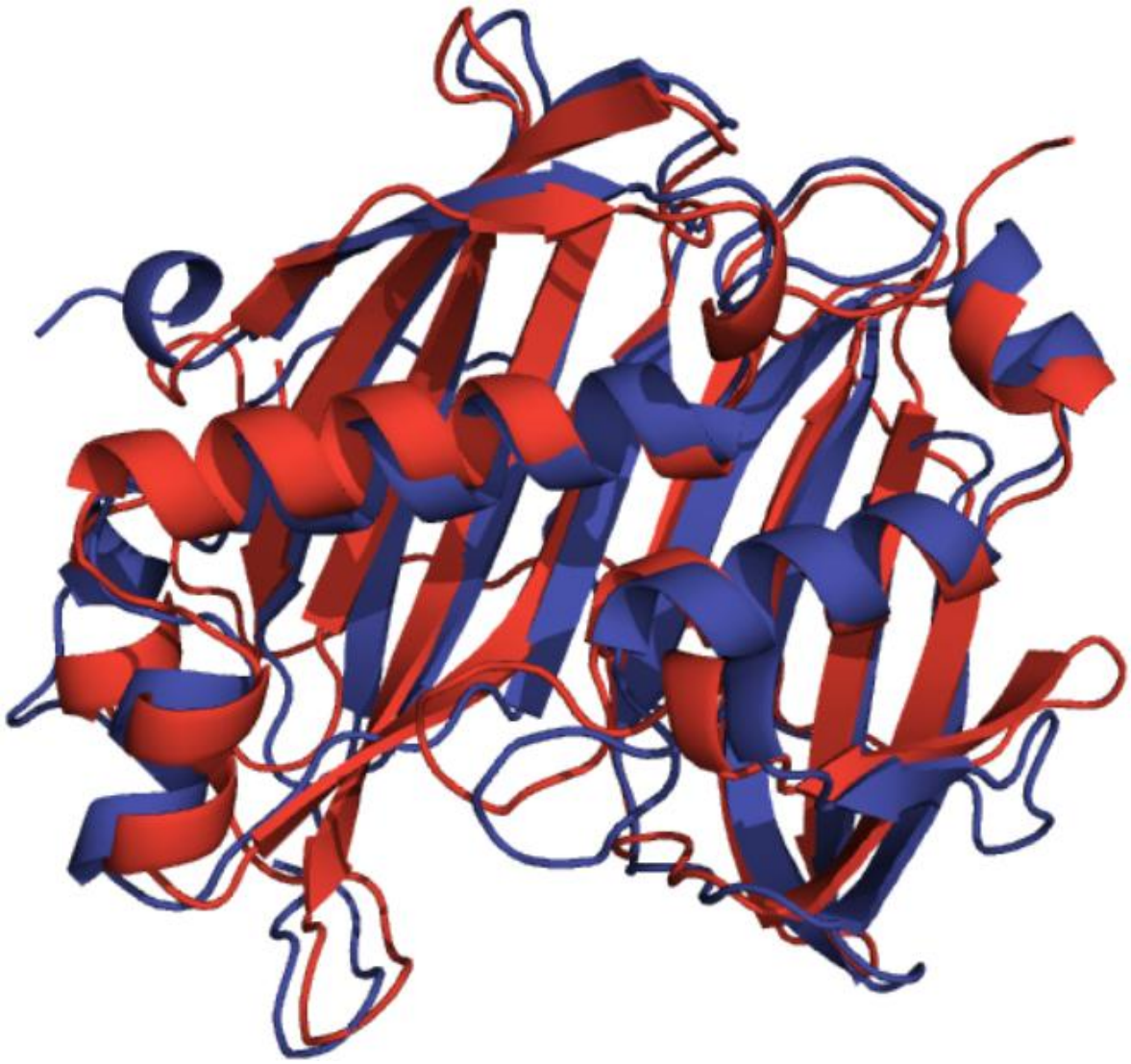


Figure 2.2 Superimposed PDB structures. 2ESS (blue) from *B. thetaioatmicron* (Subfamily F) and 2OWN (red) from *L. plantarum* (Subfamily J).

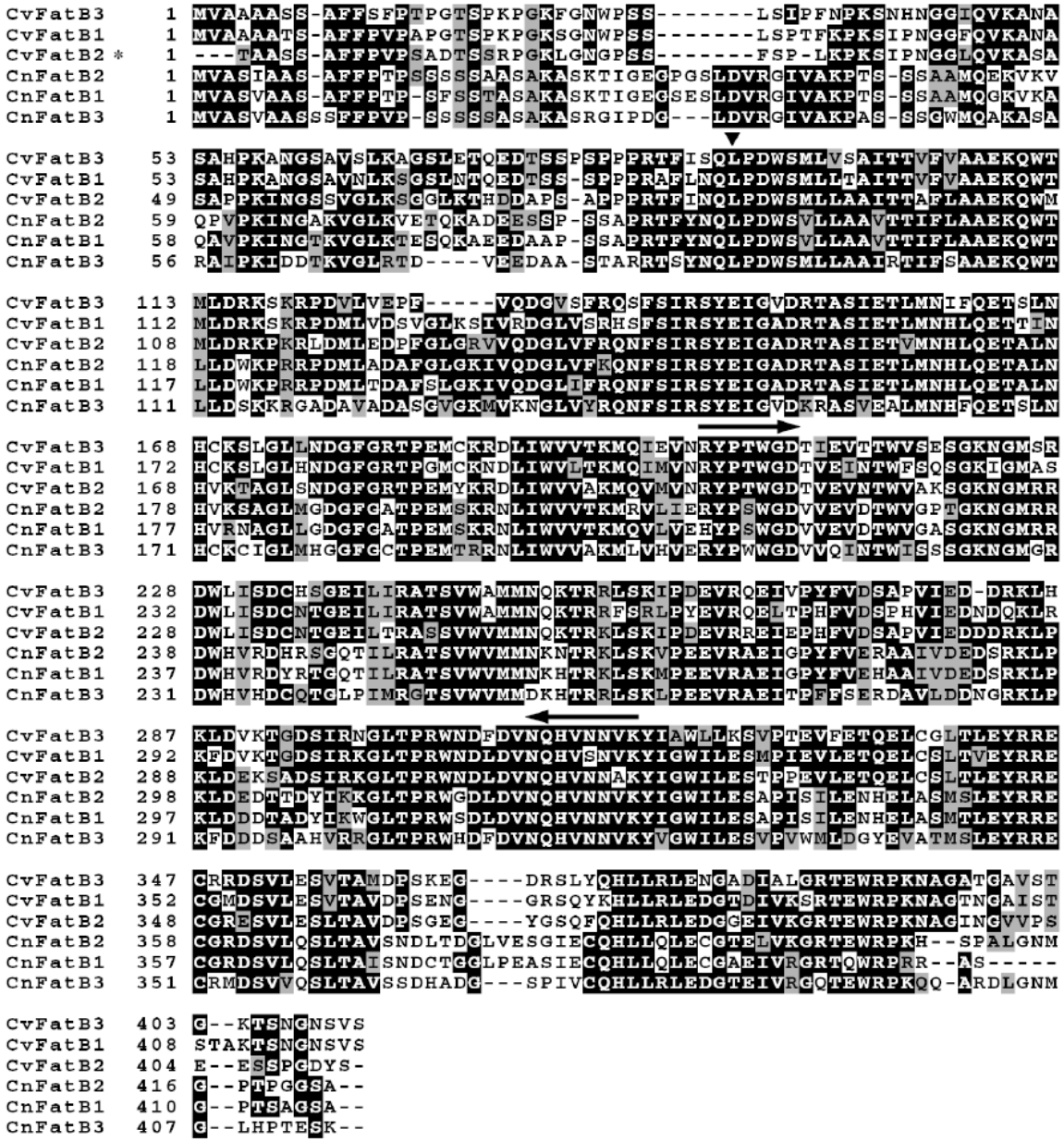


Figure 2.3 Sequence alignment of deduced amino acid sequences of *C. nucifera* (Cn) and *C. viscosissima* (Cv) acyl-ACP TEs. The putative N-terminal amino acid residue is leucine (▼). Two arrows indicate the conserved regions from which the degenerated primers were designed. The N-terminal sequence of CvFatB2 is incomplete (*).

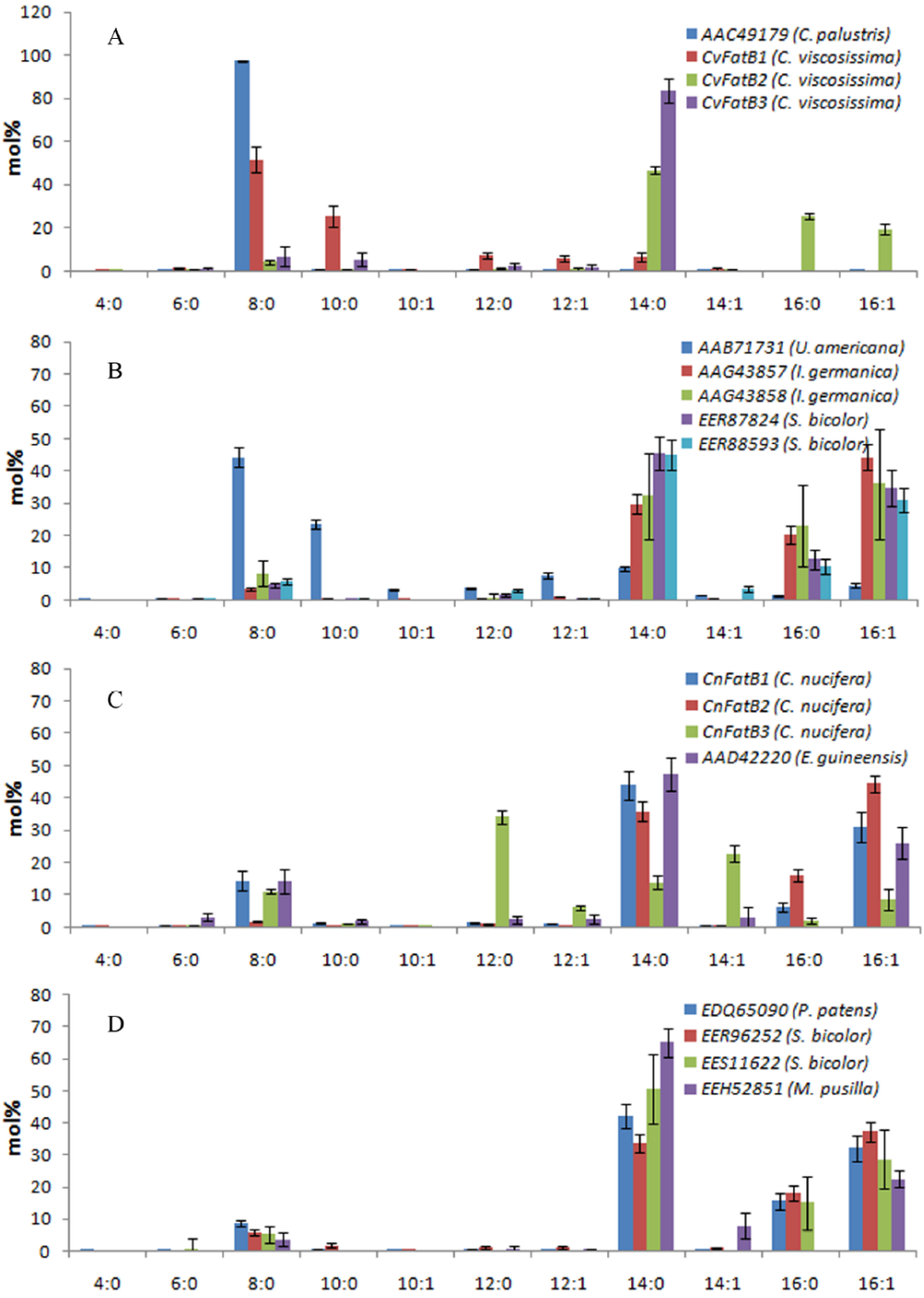


Figure 2.4 Fatty acid compositions of *E. coli* K27 cultures expressing plant acyl-ACP TEs. A: TEs from coconut and oil palm in Subfamily A; B: TEs from *C. viscosissima* and *Cuphea palustris* in Subfamily A; C: remaining TEs characterized from Subfamily A; D: TEs in Subfamily B and Subfamily D. Error bars represent standard errors.

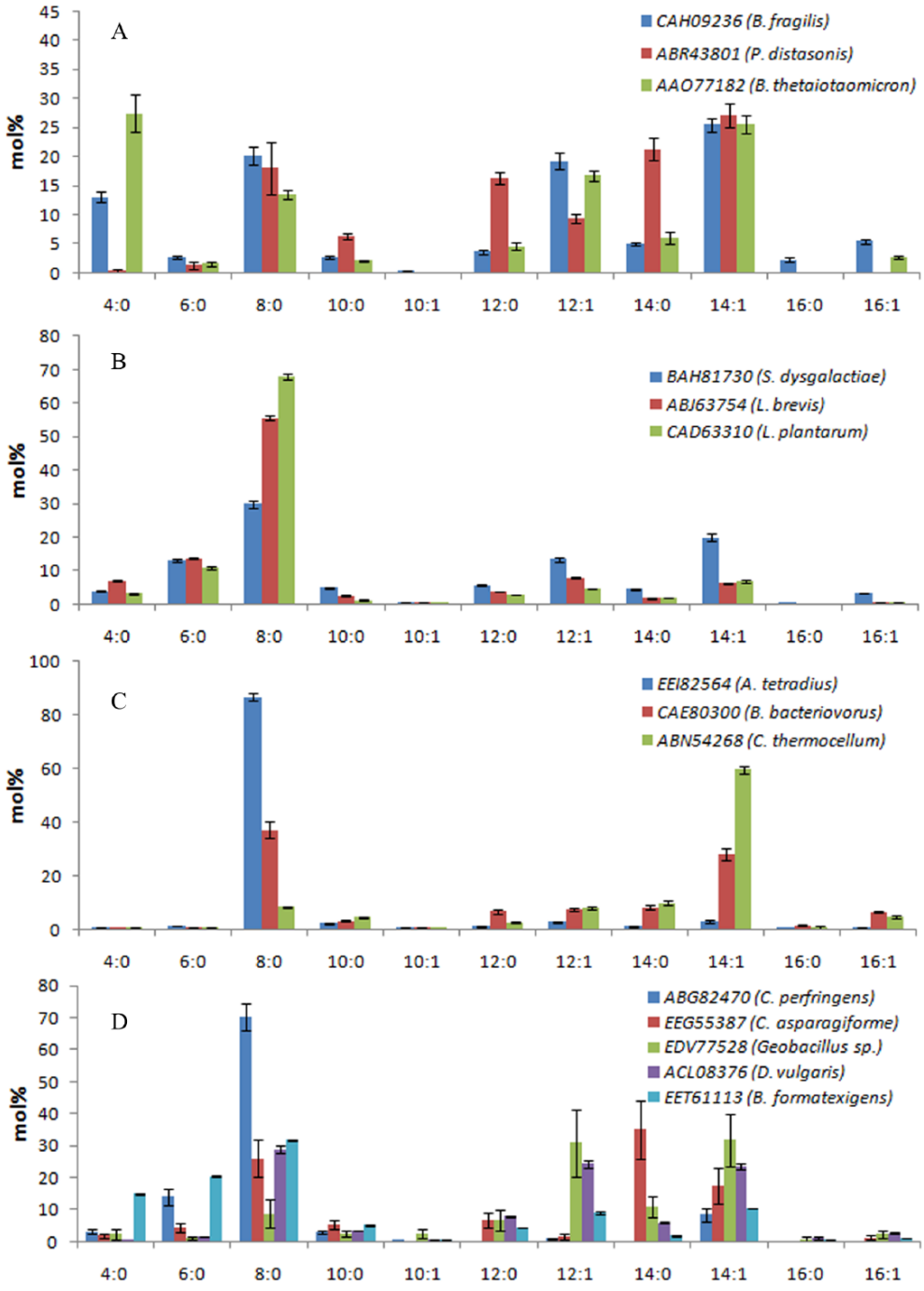


Figure 2.5 Fatty acid compositions of *E. coli* K27 cultures expressing bacterial acyl-ACP TEs. A: TEs from Subfamily F; B: TEs from Subfamily J; C: non-grouped TEs; D: other bacterial TEs. Error bars represent standard errors.

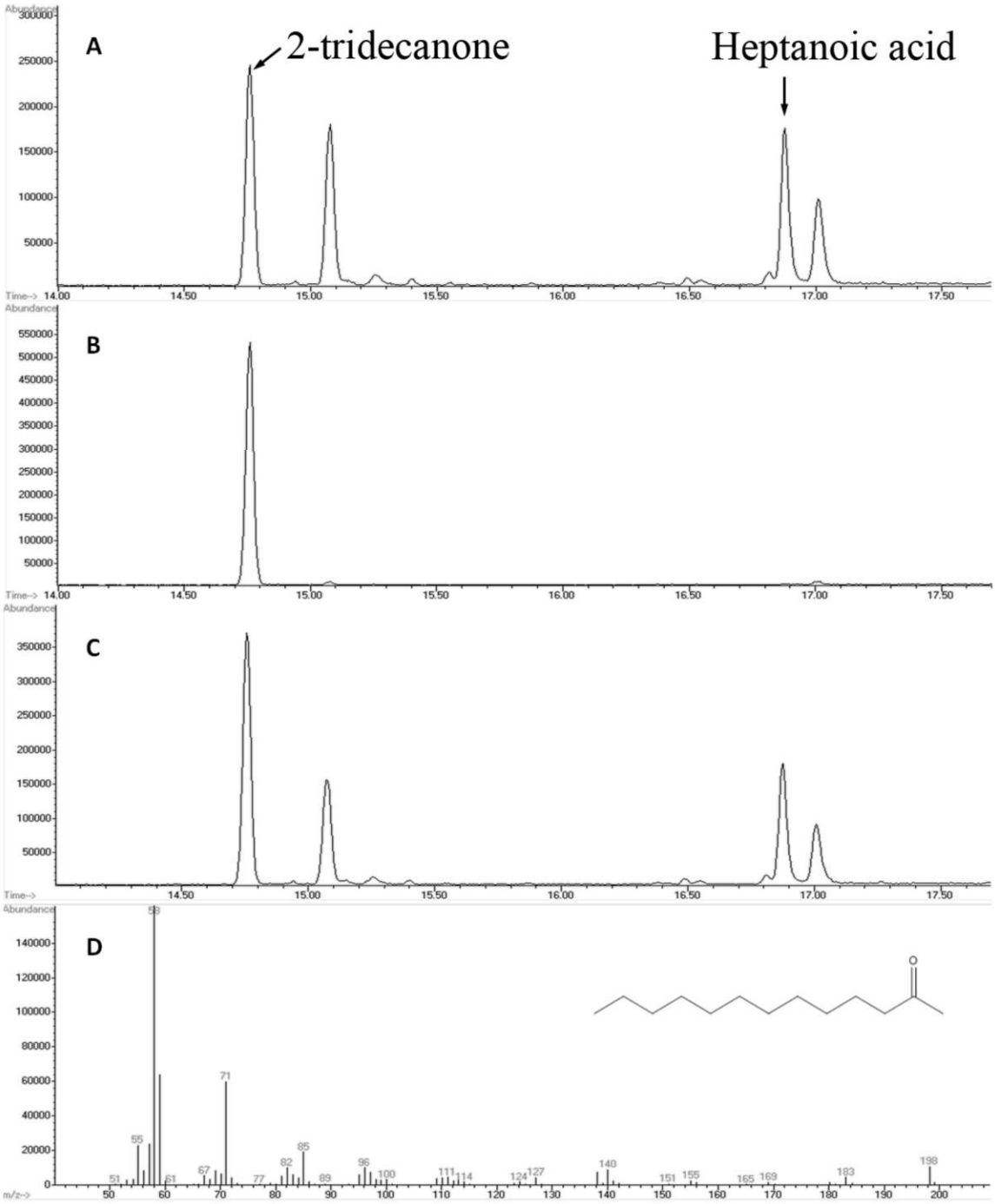


Figure 2.6 Identification of 2-tridecanone in the culture expressing a bacterial TE. A: GC of extract from *E. coli* K27 culture expressing a bacterial TE (*Bdellovibrio bacteriovorus*, GenBank:CAE80300); B: GC of 2-tridecanone standard; C: GC of a mixture of A and B; D: mass spectrum of 2-tridecanone.

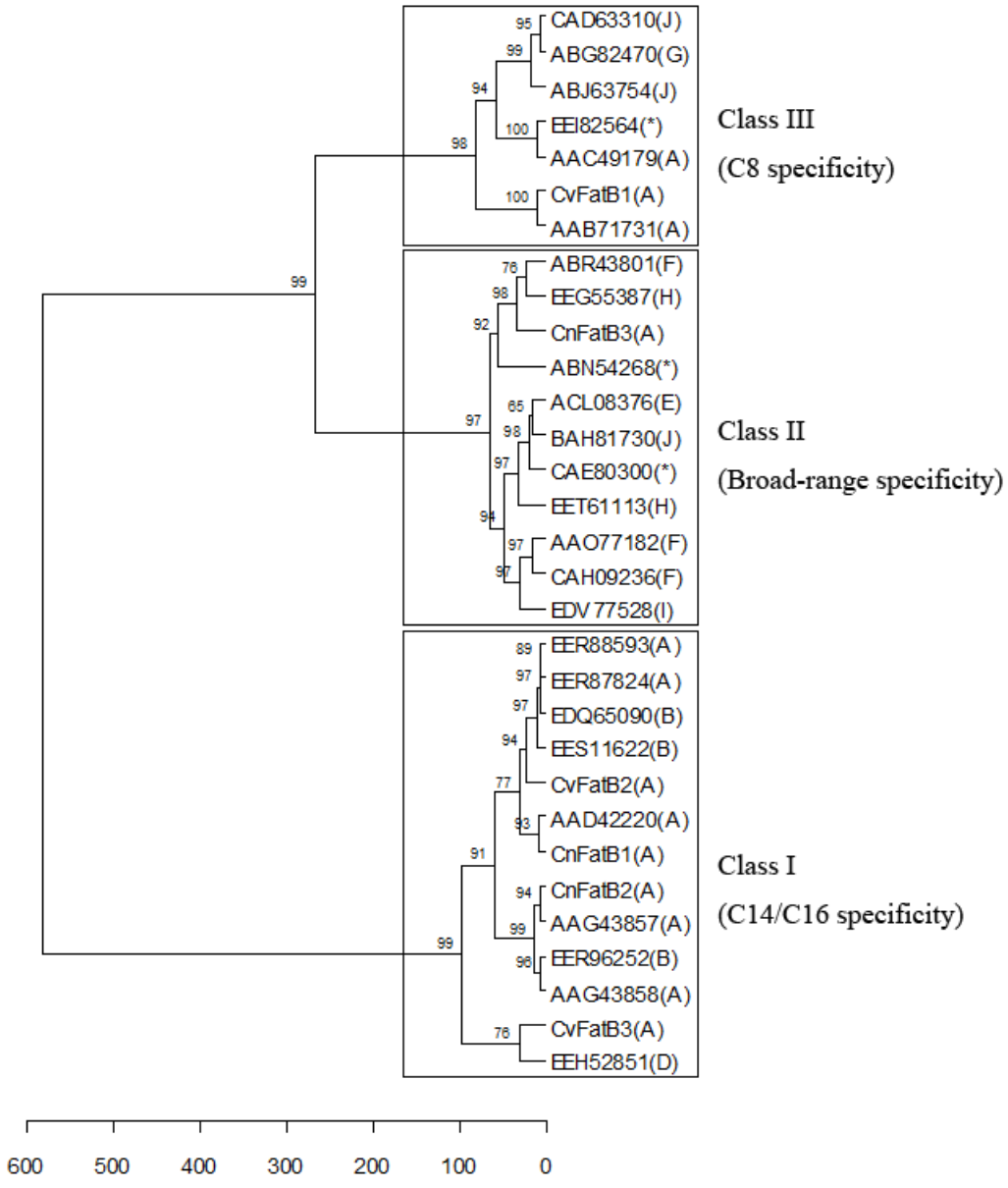


Figure 2.7 Hierarchical clustering dendrogram of acyl-ACP TEs. Cluster analysis was performed with fatty acid composition data using Euclidean distances and Ward's hierarchical clustering method. The *p*-values were calculated via multiscale bootstrap resampling with 1000 replicates. Subfamilies to which each sequence belongs are indicated in parentheses. Non-grouped sequences are indicated by asterisks.

CHAPTER 3 – IDENTIFICATION OF ACTIVE SITE RESIDUES IMPLIES A TWO-STEP CATALYTIC MECHANISM FOR ACYL-ACP THIOESTERASE

Fuyuan Jing, Marna D. Yandeau-Nelson, Basil J. Nikolau
Department of Biochemistry, Biophysics, and Molecular Biology, Iowa State University

A manuscript to be submitted for publication

Abstract

Acyl-acyl carrier protein (ACP) thioesterases (TEs), hydrolyzing the thioester bonds of acyl-ACPs to release free fatty acids, play essential roles in controlling the amount and composition of fatty acids produced by the fatty acid synthesis (FAS) pathway. Understanding the catalytic residues and mechanism will be beneficial for the engineering of this enzyme. Acyl-ACP TEs were previously proposed to utilize a papain-like protease catalytic triad consisting of Cys, His, and Asn. The increasing number of functionally characterized plant and bacterial acyl-ACP TEs and the advent of several crystal structures of bacterial acyl-ACP TEs allow us to reinvestigate the catalytic mechanism. A multiple sequence alignment of both plant and bacterial acyl-ACP TEs revealed that the previously proposed catalytic residue Cys is not conserved. Mutating Cys348 in CvFatB2 to Serine or Alanine resulted in still active mutants, demonstrating that Cys is not a catalytic residue. In contrast, the multiple sequence alignment together with the structure modeling of CvFatB2 suggested that the highly conserved Asp309 and Glu347, in addition to previously proposed Asn311 and His313, are involved in the catalysis of acyl-ACP TEs. The substantial activity loss of site-directed mutants at these positions proved their catalytic functions. By comparing the bacterial acyl-ACP TE structure (2OWN) and *Pseudomonas* 4HBT structure (1BVQ), a two-step catalytic mechanism for plant and bacterial acyl-ACP TEs was proposed.

Introduction

De novo fatty acid biosynthesis is an iterative metabolic process, commonly primed with the acetyl moiety from acetyl-CoA and extended by two carbons per cycle, via a reaction with malonyl-ACP. Plant and bacterial organisms use a type II FAS system, which integrates the capabilities of discrete enzymes that each catalyze distinct reactions in the fatty acid elongation cycle (White et al. 2005). Acyl-acyl carrier protein (ACP) thioesterase (TE) plays an essential role in the termination of acyl chain elongation in these *de novo* fatty acid synthesis systems (Jones et al. 1995; Voelker 1996) by catalyzing the hydrolysis of the thioester bond of the acyl-ACP to release the free fatty acid product of the system. These acyl-ACP TEs can exhibit diverse substrate specificities, and terminate fatty acid biosynthesis to produce saturated and unsaturated fatty acids ranging from 4- to 18-carbons in chain length (Jing et al. 2011). Because of the important role in determining fatty acid chain length, acyl-ACP TEs have been widely studied and used to modify the fatty acid composition of oil crops for improved nutritional value or industrial uses (Ohlrogge 1994; Voelker et al. 1996; Hawkins et al. 1998; Hills 1999).

More recently, these enzymes have been the targets of microbial metabolic engineering for the production of biorenewable chemicals (Zhang et al. 2011; Lennen et al. 2012). This application has increased the demand for specific fatty acids across a broad range of chemical functionalities. Therefore, the better understanding of the catalytic mechanism of acyl-ACP TEs would be beneficial for the targeted engineering of a more efficient acyl-ACP TE. Although substrate specificities of many acyl-ACP TEs have been functionally characterized (Jing et al. 2011), the catalytic residues and catalytic mechanisms for these enzymes are still not well understood (Mayer et al. 2005). As an essential enzyme that controls the amount and composition of fatty acid output from one of the most import metabolic pathways, there is fundamental significance to studing the catalytic mechanism of acyl-ACP TE as part of an effort to understand the regulation of FAS pathway.

Several efforts have been made to identify the catalytic residues and understand the catalytic mechanism shared by acyl-ACP TEs, specifically from plants. The sensitivity of plant acyl-ACP TEs to thiol inhibitors suggested that these enzymes employ a cysteine in catalysis (Davies et al. 1991; Pollard et al. 1991). A conserved cysteine and histidine were proposed as catalytic residues based on a site-directed mutagenesis study (Yuan et al. 1996). A third catalytic residue, Asn was identified later based on structure modeling and mutation results (Mayer et al. 2005). From these results, a papain-like catalytic triad containing the amino acid residues, Asn, His, and Cys, was proposed for plant acyl-ACP TEs.

Plant and bacterial acyl-ACP TEs have been proposed to exhibit the same "hotdog" tertiary structure, even though they share little sequence similarity (Cantu et al. 2010). No plant acyl-ACP TE structure has ever been experimentally determined, but several bacterial TE structures are available in the PDB database (2ESS, 2OWN, 4GAK). In the past decade, the number of acyl-ACP TE sequences in the public databases has exponentially increased, with the majority being from bacterial sources. Also, an increasing number of plant and bacterial acyl-ACP TEs have been functionally characterized (Jing et al. 2011). With the availability of these additional sequences and structural information, in this study we reinvestigated the catalytic residues of acyl-ACP TE. Multiple sequence alignment of plant and bacterial acyl-ACP TE sequences, together with structure prediction models were used to predict the catalytic residues for acyl-ACP TEs. Site-directed mutagenesis has demonstrated that Asp309 and Glu347 in *Cuphea viscosissima* acyl-ACP TE (CvFatB2; GenBank Accession AEM72523.1) are involved in catalysis, instead of the previously proposed Cys348 (Yuan et al. 1996). Based on the identification of these additional catalytic residues, a two-step catalytic mechanism is being proposed for the acyl-ACP TE catalyzed reaction.

Methods

Computational modeling and predictions

A multiple sequence alignment of 28 functionally characterized acyl-ACP TEs (Jing et al. 2011), including 16 plant and 12 bacterial sequences, was generated with ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) using the default parameters.

The secondary structure of CvFatB2 was predicted using the PSIPRED Protein Structure Prediction Server (http://bioinf.cs.ucl.ac.uk/psipred/) (Jones 1999). The tertiary structure of CvFatB2 was predicted using I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) with default parameters (Zhang 2008), and was viewed and analyzed with the PyMOL Molecular Graphics System, Version 1.5.0.4 (Schrödinger, LLC).

Site-directed mutagenesis

Three acyl-ACP TE coding cDNAs were previously cloned from *Cocos nucifera* (CnFatB2) and *Cuphea viscosissima* (CvFatB1, and CvFatB2) (Jing et al. 2011). The cDNA encoding the mature part of each protein (i.e. without the N-terminal chloroplast transit peptide sequence) was codon-optimized for expression in *E. coli*, chemically synthesized and cloned into a pUC57 vector such that it was transcriptionally driven by the *lacZ* promoter (Jing et al. 2011). Using these plasmids, mutations at specific conserved sites were generated with the QuikChange site-directed mutagenesis kit (Agilent Technologies, USA), following the manufacturer's instructions. Residues targeted for mutations were Asn321, His323, Glu357, and Cys358 in CnFatB2; Asn315, His317, Glu351, and Cys352 in CvFatB1; and Asp309, Asn311, His313, Glu347, and Cys348 in CvFatB2. For each acyl-ACP TE, Cys was mutated to Ser or Ala, Glu was mutated to Asp, Gln, or Ala, and both Asn and His were mutated to Ala. The Asp309 of CvFatB2 was mutated to Ala, Asn, and Glu. Authenticity of all mutants was confirmed by sequencing both strands of all constructs.

In vivo activities of acyl-ACP TE mutants

E. coli strain K27 (*fadD88*) contains a mutation in the *fadD* gene, which impairs β-oxidation of fatty acids and results in the accumulation of free fatty acids in the growth medium. Each acyl-ACP TE was expressed in the K27 strain, and free fatty acids that accumulated in the medium were extracted and analyzed (Jing et al. 2011). Fatty acid extracts from four independent colony isolates for each construct were independently analyzed via GC-MS. The total concentration of fatty acids produced by each acyl-ACP TE was obtained by subtracting the concentration of fatty acids produced by *E. coli* expressing a control plasmid (pUC57) that lacked an exogenous acyl-ACP TE.

Determination of acyl-ACP TE protein expression level

Using purified CvFatB2 protein, polyclonal antibodies were produced in a mouse host at the Hybridoma Facility of Iowa State University (http://www.biotech.iastate.edu/facilities/hybridoma/). To assess the acyl-ACP TE protein expression level, soluble proteins were extracted from microbial cell pellets and analyzed by immuno-blotting. Cell pellets were resuspended and incubated for 30 min in Lysis Buffer (50 mM Tris-HCl pH8.0, 150 mM NaCl, 10% glycerol, 0.6 mM phenylmethanesulfonylfluoride, and 0.2 mg/ml lysozyme). Following sonication for 5 s, the suspension was subjected to centrifugation at 13,000 x g for 5 min, and the soluble protein fraction was recovered in the supernatant. Protein concentration was quantified using the Bio-Rad Protein Assay Kit (Bio-Rad, USA). The expression level of the wild type and point-mutant derivatives of CvFatB2 were quantified by first subjecting 35 μg of soluble crude protein extract to SDS-PAGE in a 12% polyacrylamide gel, and transferring the separated proteins to nitrocellulose membrane. Membranes were reacted with the primary CvFatB2 antibody at 1:2000 dilution, and the secondary antibody Goat-Anti-Mouse IgG (H + L)-HRP (Bio-Rad, USA) at 1:3000 dilution. Immuno-detection was performed using an enhanced chemifluorescence western-blotting detection kit according to the manufacturer's instructions (Thermo Scientific, USA) and visualized using a ChemiDoc™ XRS+ System (Bio-Rad, USA). Images were analyzed with Image

Lab™ Software (Bio-Rad, USA). The protein expression level for CvFatB2 in *E. coli* strain K27 was defined as 100% and the relative expression level of each mutant protein was calculated using the ratio of the band intensity of each mutant to the wild-type protein.

Purification of acyl-ACP TE mutant proteins

Each acyl-ACP TE was expressed in *E. coli* using a modified pET30b(+) vector, named pET30f. In this modified vector, expression was under the control of the T7 promoter, but the original His-tag, thrombin protease site, S-tag, and enterokinase protease sites present in pET30b(+) were replaced by a His-tag and a TEV protease recognition site. Each expression construct was confirmed by sequencing both strands of the vectors. Expression was in the host, *E. coli* BL21 Star (DE3) (Life Technologies, USA). High level expression of acyl-ACP TEs was initiated from a single colony inoculation of an overnight 5 ml of culture of LB medium supplemented with 50 mg/L kanamycin. This culture was used to inoculate a 1 L LB medium supplemented with 50 mg/L kanamycin, and grown at 37℃. When the OD600 reached 0.5-0.7, isopropyl-β-D-galactopyranoside (IPTG) was added to a final concentration of 0.4 mM, and the incubation continued at 25℃ for 16-20 h. The cells were harvested by centrifugation, and resuspended with Lysis Buffer containing 20 mM imidazole, followed by centrifugation at 10,000 x g for 30 min. The supernatant was applied to a Ni-NTA column, pre-equilibrated with buffer A (50 mM Tris-HCl pH8.0, 150 mM NaCl, 10% glycerol) containing 20 mM imidazole. After successive washing of the column with buffer A supplemented with 20 mM, 40 mM, and 60 mM imidazole, the acyl-ACP TE protein was eluted with 250 mM imidazole in buffer A. An Amicon Ultra 10 kDa centrifugal filter (EMD Millipore, USA) was used for buffer exchange and concentrating each protein preparation.

CD spectra of acyl-ACP TE proteins

To assess the extent of folding of each purified protein, 0.2 mg/ml of protein, dissolved in 2 mM sodium phosphate buffer, pH8.0, was subjected to far-UV CD spectrum analysis using a Jasco J-710 spectropolarimeter (Jasco, USA) at 25℃. The cell path length was 0.1 cm, and the spectra were an average of two scans collected at a speed of 50 nm/min and a step resolution of 0.2 nm. Protein secondary structure was predicted from CD spectra using JFIT software (B. Rupp; http://www.ruppweb.org/cd/cdtutorial.htm).

Results

Sequence alignment analysis of acyl-ACP thioesterases

Twenty-eight acyl-ACP TEs, including 16 plant TEs and 12 bacterial TEs, have previously been functionally characterized (Jing et al. 2011). Multiple sequence alignment of these acyl-ACP TEs established that they exhibit 44% overall amino acid sequence similarity and only 1.2% overall identity. However, the region from amino acid positions 301 to 350 of CvFatB2, defined as the catalytic domain, shows a higher conservation among these 28 proteins (86% similarity and 10% identity) (Figure 3.1). This region is even more highly conserved among the 16 plant TEs (~36% identity). These sequence alignments revealed 6 residues that are absolutely conserved across the entire TE sequence, and these are Trp192, Asp309, Asn311, His313, Tyr319, and Glu347 (numbered relative to the sequence of CvFatB2) (Figure 3.1, Figure S3.1). Among these 6 conserved residues, five are located within the smaller catalytic domain (Figure 3.1), which also contains the previously proposed active-site motifs of N311Q(K)HVN(S)N and Y343RR(K)EC(Q/T) of plant acyl-ACP TEs (Yuan et al. 1996). The first motif is conserved among all plant and bacterial sequences, while the second motif is conserved only in plant sequences.

The catalytic residues of acyl-ACP TEs were previously proposed to be Asn311, His313, and Cys348 (numbered relative to the sequence of CvFatB2) (Mayer et al. 2005).

However, when one compares the sequences of all 28 TE sequences in Figure 3.1, the Cys348 residue that is located in the second proposed active-site motif was only conserved across plant TEs, whereas the Asn311 and His313 residues that are located in the first proposed active-site motif are conserved among all 28 TEs aligned in Figure 3.1. Furthermore, no corresponding Cysteine residue in this region could be identified in any of the bacterial TE sequences analyzed herein. Interestingly, residue Glu347, which is located in the second proposed active-site motif immediately adjacent to, and upstream of the proposed Cys catalytic-residue, is conserved among all 28 TE sequences.

To further investigate the level of conservation of these potential catalytic residues (i.e. Trp192, Asp309, Asn311, His313, Tyr319, Glu347, and Cys348) among a wider collection of TEs, a total of 1019 sequences (collected on 10/15/2012) were collected from the ThYme database (Cantu et al. 2010) and these were also aligned. Residues Trp192, Asp309, Asn311, His313, Tyr319, Glu347, and Cys348 (positions relative to CvFatB2) are conserved among 95.8%, 99.8%, 93.1%, 99.9%, 98.7%, 85.9%, and 21.8% of the 1019 sequences, respectively. We conclude therefore that the highly conserved residues (Trp192, Asp309, Asn311, His313, Tyr319, and Glu347) may play important structural or catalytic roles in TE functionality. The previously proposed catalytic residue, Cys348 is not as highly conserved, suggesting that its role in catalysis may not be direct.

Predicted structure of a plant acyl-ACP TE

Using the I-Tasser algorithm (Zhang 2008), we modeled the three-dimensional structure of *C. viscosissima* C14/16-ACP TE CvFatB2 using as template the crystallographically-determined TE structure, 2OWN from *Lactobacillus plantarum*. The resulting predicted structure is similar to a previous structural prediction of plant acyl-ACP TEs modeled using the more distant *E. coli* TEII (PDB IC8U) (Mayer et al. 2005). The predicted structure of CvFatB2 is comprised of two helix/4-stranded sheet domains in the canonical “hotdog” fold (Mayer et al. 2005; Cantu et al. 2010). Two hotdog domains are linked by a long coil structure, and the two proposed active-site

motifs, (N311Q(K)HVN(S)N and Y343RR(K)EC(Q/T)) are located in the C-terminal hotdog structure on the edge of the cleft between two hotdog domains.

The six highly conserved residues (Trp192, Asp309, Asn311, His313, Tyr319, and Glu347) identified in the multiple sequence alignment are similarly oriented in the *L. plantarum* TE structure and in the homology model of CvFatB2 (Figure 3.2). One of these, Trp192 in CvFatB2, resides in the N-terminal hotdog domain that is far removed from the active site, indicating that this residue may be structurally important, but not catalytically. The other conserved residues (Asp309, Asn311, His313, Tyr319, and Glu347 in CvFatB2) are in the C-terminal hotdog domain. Among these residues, Tyr319 is located in the central α-helix and is distant from the proposed active site, indicating that it may also be only of structural importance. In the predicted CvFatB2 structure (Figure 3.2), the previously proposed Cys348 catalytic residue (Yuan et al. 1996) is oriented distantly from the proposed catalytic residues, His313 and Asn311. Moreover, no analogous Cys residue can be identified that resides nearby His177 and Asn175 in the 2OWN tertiary structure from the *L. plantarum* TE (equivalent to H313 and Asn311 in the CvFatB2 sequence). Together, these data suggest that Cys348 is not involved in TE catalysis. In contrast, residues Asp309 and Glu347, together with the previously proposed catalytic residues Asn311 and His313 are on a coil structure and cluster in close proximity to each other. We therefore suggest that these latter four residues may be more directly involved in TE catalysis.

Mutagenesis-based evaluation of potential catalytic residues of acyl-ACP TE

To verify the prediction that Asp309, Glu347, Asn311 and His313 are catalytic residues, and that Cys348 is not involved in catalysis, site-directed mutants at these positions were generated for the *C. viscosissima* TE CvFatB2. The *in vivo* activity of each mutant was determined by expression in *E. coli* strain K27, and the free fatty acids that accumulated in the media were quantified as a representation of the catalytic activity of each TE. We initially evaluated mutants of Cys348, which prior studies have suggested may be part of the catalytic triad that catalyzes the hydrolysis of the thioester

bond in acyl-ACP (Yuan et al. 1996). Replacing this Cys residue with either Ser or Ala (i.e. CvFatB2-C348S and C348A) did not result in inactive mutants. CvFatB2-C348A exhibited ~60% *in vivo* activity of wild-type, and C348S exhibited ~130% *in vivo* activity of the wild-type (Figure 3.3A). In contrast, all mutants of Asp309, Asn311, His313, and Glu347 had dramatically decreased *in vivo* activity.

To ensure that the differences in fatty acid yields generated by the wild-type and mutant variants of CvFatB2 TEs can be attributed to the individual mutations that disrupt active site function and not to differences in levels of expression of the heterologous TEs, we evaluated the protein expression level of all CvFatB2 TEs in *E. coli* strain K27. Specifically, crude extracts from every expressing strain was assessed via western blot analysis with the anti-CvFatB2 antibody. The expression of mutant CvFatB2-C348S was ~65% of the wild-type, while the expression of all the other mutants was higher than the wild-type TE, in the range of 130% to 700% of the wild-type levels (Figure 3.3B). Total fatty acid accumulation for CvFatB2 and its mutants was thus normalized for this difference in the protein expression level. The normalized activity of C348A and C348S was 50% and 200% of the wild-type, respectively. Similarly, mutations of Asp309, Asn311, His313 and Glu347 were evaluated, and all these resulted in the reduction of relative TE activity to less than 1% of wild-type, whereas protein expression for each mutant was at or higher than wild-type levels (Figure 3.3B). Together, these data demonstrate that reduction of *in vivo* activities of the mutants is not contributed to different expression levels of mutant TE proteins, but these are due to the specific alteration of the individual residues.

To assess whether the active site mutations in CvFatB2 resulted in altered protein folding properties, far-UV CD spectra of CvFatB2 wild-type and all purified mutant proteins were obtained. Comparing the spectrum of each mutant protein to the wild-type protein indicated no significant differences in secondary structure between mutants and wild-type (Figure 3.4). The secondary structures of CvFatB2 and its mutants were calculated using the program, Jfit. CvFatB2 protein exhibited 11.9% α-helix, 24.0% β-sheet and 64.1% coil structures, and each mutant protein shared a similar secondary

structure profile (Table S3.1) that did not differ significantly from wild-type (*p*-value>0.1). Moreover, a computational prediction of the secondary structure of CvFatB2 using the PSIPRED Protein Structure Prediction Server produced a similar secondary structure composition (11.9% α-helix, 28.5% β-sheet, and 59.6% coil) to that calculated from CD spectra (*p*-value=0.15). All these results suggest that the activity loss of Asp309, Asn311, His313, and Glu347 mutants was not caused by incorrect folding of mutant proteins, but due instead to the specific changes of these active site residues and subsequent loss of catalytic activity.

The role of these specific residues in TE catalysis was further evaluated by mutating the equivalent residues in an additional two acyl-ACP TEs templates, including *C. nucifera* TE CnFatB2 and *C. viscosissima* TE CvFatB1. Prior analyses of these TEs indicated that they each shows different substrate specificities, with CnFatB2 hydrolyzing acyl-ACPs of 14- and 16-carbon fatty acids, and CvFatB1 hydrolyzing acyl-ACPs of 8- and 10-carbon fatty acids (Jing et al. 2011). Replacing the Cys residue previously proposed as having a catalytic role (i.e. Cys358 in CnFatB2 and Cys352 in CvFatB1) with either Ser or Ala in both acyl-ACP TEs resulted in only slight change of *in vivo* activity, as compared to the wild-type enzymes (Figure 3.5). In contrast, mutations of other residues that we propose to be involved in catalysis (i.e. Asn321, His323, and Glu357 in CnFatB2 and Asn315, His317, and Glu351 in CvFatB1) caused dramatic decrease in the *in vivo* activity of these enzymes (Figure 3.5). These results from mutagenesis of acyl-ACP TEs with different substrate specificities (CnFatB2 and CvFatB1) are consistent with observation in CvFatB2 mutants, supporting our conclusion that the Cys is not directly involved in catalysis of acyl-ACP TE and the adjacent Glu may be one of the catalytic residues.

Discussion

The Asn-His-Cys catalytic triad model of the acyl-ACP TE catalytic mechanism was proposed at a time when only a small number of functionally characterized acyl-ACP TE sequences, primarily from plants, were available for a sequence-homology

based prediction (Yuan et al. 1996). In the past decade, a large number of additional acyl-ACP TEs have been isolated and characterized from diverse plants and bacteria (Jing et al. 2011) and high-throughput sequencing projects have generated numerous putative acyl-ACP TE sequences from both plants and bacteria. In addition, the last decade has seen the advent of several crystal structures of bacterial acyl-ACP TEs. With this abundance of new sequence and structural data comes an increased ability to computationally predict the catalytic residues in acyl-ACP TEs, based upon both primary and tertiary structural modeling.

Cys348 is not a catalytic residue

Because both plants and bacteria utilize a type II fatty acid synthase system (White et al. 2005) it is reasonable to postulate that acyl-ACP TEs in these organisms are evolutionarily related proteins, and thus utilize the same catalytic mechanism. An indicator of the latter would be the finding that the catalytic residues have been conserved among these enzymes during evolution. Beginning with this premise we assembled a multiple sequence alignment of over 1,000 sequences of both plant and bacterial acyl-ACP TEs, and found the previously proposed catalytic Cys residue (Cys348 in CvFatB2) is poorly conserved among these TEs, and may therefore not be part of the hydrolysis catalytic mechanism.

Replacing Cys348 with Ser resulted in a mutant that was still very active compare to the wild-type enzyme, which is consistent with a previous report (Yuan et al. 1996). This result is expected regardless of whether Cys acts as a catalytic residue, due to the ability of Ser to act as a nucleophile, as it does in serine proteases (Hedstrom 2002). When Cys348 was mutated to an Ala, which would have no catalytic function, the mutant retained more than 50% activity, demonstrating that the Cys residue is not a member of the catalytic triad. This result, however, is in contrast to a previous report, where mutating Cys to Ala in a 12:0-ACP TE UcFatB1 totally inactivated the TE enzyme (Yuan et al. 1996). In this study, we analyzed Cys-to-Ala mutants in CnFatB2, CvFatB1 and CvFatB2 and consistently observed that the mutants retained high levels of

activity. In combination, these data demonstrated that acyl-ACP TEs do not use a papain-like protease mechanism.

Cys348 was previously proposed as a catalytic residue, in part, based on the observation that plant acyl-ACP TEs displayed sensitivity to thiol inhibitors and Cys to Ser mutation could convert the sensitivity to serine hydroxyl-reactive reagents (Davies et al. 1991; Pollard et al. 1991). For example, 0.1 mM 5,5'-dithiobis(2-nitrobenzoic acid) inhibited 92% of TE activity, 1 mM iodoacetamide inhibited 72% of TE activity, and 5 mM phenylmethylsulfonyl fluoride showed 90% inhibition effect on serine mutant (Yuan et al. 1996). As reagents 5,5'-dithiobis(2-nitrobenzoic acid) and iodoacetamide bind covalently with the thiol group of cysteine and phenylmethylsulfonyl fluoride binds covalently with the hydroxyl group of serine, these inhibitory effects were originally considered as an indication that Cys348 is a catalytic residue. However, with the discovery that the adjacent residue Glu347 instead of Cys348 is the catalytic residue, there is another explanation for those inhibition effects: the covalently attached thiol inhibitors may force a conformational change of the adjacent catalytic residue, Glu347. Alternatively, the inhibitor might block the access of substrates into the active site, and thus inhibit the TE activity.

Identification of Asp309 and Glu347 as catalytic residues

Past studies of the catalytic mechanism of acyl-ACP TEs have focused primarily on plant sourced enzymes, and these efforts have been impeded by the lack of a 3-D structure of such an enzyme. Recently, the use of homology modeling of proteins has become more accurate in their predictive power (Ginalski 2006). In this study, the structure of CvFatB2 was predicted using the structure of an acyl-ACP TE (2OWN) from the bacterium, *L. plantarum*. The predicted CvFatB2 structure is consistent with a previous structural model of Arabidopsis AtFatB (Mayer et al. 2005). Integrating this tertiary structure prediction, with multiple sequence alignment of a wider phylogenetic range of TEs, we identified a set of residues (Asp309, Asn311, His313, and Glu347) that are highly conserved in over 1000 TE sequences, and are predicted to be in close

proximity in 3-D space, which may suggest a role for these residues in the catalytic mechanism.

These predictions were tested by site-directed mutagenesis of Asp309, Glu347, Asn311 and His313 of CvFatB2. Mutating the residue Glu347 to either the similar acidic residue Asp or the non-active residue Ala resulted in almost complete loss of enzymatic activity, suggesting that Glu347 is a catalytic residue. Although Asp shares similar chemical properties with Glu, it is smaller in size and the Glu-to-Asp mutant does not support catalytic activity. This suggests that the proximity of Glu347 to other catalytic residues or substrates is necessary for catalysis. Moreover, mutation of the Glu to Gln in CvFatB2 resulted in substantial loss of activity, suggesting that the negative charge of the carboxylic group of Glu347 is also important. Bioinformatic analysis also suggested a role for Asp309 in catalysis. Similar to Glu347, mutating residue Asp309 to Ala nearly inactivated the enzyme. Moreover, mutation to the larger residue, Glu, or the uncharged residue, Asn, also reduced enzymatic activity. Collectively, these data indicate that residue Asp309 is a catalytic residue and the charge and position of the carboxylic group may play an important role in the catalysis. Residues Asn311 and His313 were also mutated to Ala in our study. As expected, only negligible activities were detected. Mutation of the corresponding residues in two other acyl-ACP TEs (CnFatB2 and CvFatB1) resulted in the same results, further supporting the catalytic function of these residues. Interestingly, mutating His313 to Ala resulted in the lowest activity among all mutants analyzed. This may suggest that His313 plays a key function within the catalytic site.

A proposed catalytic mechanism for acyl-ACP TEs

The mutational analyses suggest that Asp309, Asn311, His313, and Glu347 are catalytic residues and therefore may function via a different catalytic mechanism. Thioesterase families, including acyl-CoA TE and acyl-ACP TE, mainly have two different folds: the hotdog fold and α/β-hydrolase fold (Cantu et al. 2010). Whereas the Ser-His-Asp catalytic triad is highly conserved among TEs having the α/β-hydrolase-

fold, hotdog-fold TEs have a variety of catalytic residues and mechanisms (Cantu et al. 2010). After reviewing the catalytic mechanisms reported for several hotdog-fold TEs (Pidugu et al. 2009; Cantu et al. 2010), two TEs that catalyze the same reaction but belong to two distinct clades of families were of interest: a 4-hydroxybenzoyl-CoA TE (4HBT) from *Pseudomonas* sp. Strain CBS3 and another 4HBT from *Arthrobacter* sp. strain SU (Thoden et al. 2002; Thoden et al. 2003). The *Pseudomonas* 4HBT contains a two residue catalytic site, utilizing a Tyr and an Asp. Catalytic residue Tyr24 is located at the N-terminus of the α-helix and is responsible for the polarization of the thioester carbonyl group via hydrogen bonding with its backbone amide, NH. Catalytic residue Asp17 acts as a nucleophile to attack the thioester carbonyl carbon from an anhydride intermediate, which then undergoes hydrolytic cleavage at the hydroxylbenzoyl carbonyl carbon atom (Zhuang et al. 2012). In comparison, *Arthrobacter* 4HBT achieves catalysis via a four-residue mechanism, utilizing a Glu, Gly, Gln and Tyr. Glu73 functions as a nucleophilic catalyst, Gly65 and Gln58 stabilize the thioester moiety via hydrogen bonding, and Tyr77 orients the water nucleophile for attack at the carbonyl carbon of the enzyme-anhydride intermediate (Song et al. 2012). Although *Pseudomonas* 4HBT and *Arthrobacter* 4HBT have different catalytic residues and different active site structures, their catalytic mechanisms share some similarities: (1) both have an acidic residue as the nucleophilic catalyst; (2) the backbone amide NH group of a residue at the N-terminus of the α-helix (Tyr24 in *Pseudomonas* 4HBT and Gly65 in *Arthrobacter* 4HBT) can form hydrogen bond with the carbonyl group of the thioester to make the carbonyl carbon more susceptible to nucleophilic attack; (3) both active sites can form an enzyme-anhydride intermediate. Comparison of the active site structure of acyl-ACP TE (2OWN) with *Pseudomonas* 4HBT (1BVQ) and *Arthrobacter* 4HBT (1Q4S) reveals that the catalytic residue Asp173 in 2OWN has the same position and orientation as the catalytic residue Asp17 in 1BVQ and they are located in well aligned loop structures (Figure 3.6A and B). Moreover, the backbone amide NH group of Asn180 in 2OWN also perfectly aligns with the NH of catalytic residue Tyr24 in 1BVQ, though the side chains of these two residues lie in opposite directions. In a multiple sequence alignment of 1019 TE

sequences, the residue Asn180 was also found to be highly conserved. Based on our results and the comparison of acyl-ACP TE structure with two 4HBT, we propose a two-step catalytic mechanism for plant and bacterial acyl-ACP TEs. In the first step, similar to the mechanism of 4HBT, the thioester carbonyl group of the acyl-ACP substrate is stabilized via hydrogen bonding with backbone amide NH group of residue Asn180, and Asp173 initiates nucleophilic attack to the carbonyl carbon of the thioester to form an anhydride intermediate. In the second step, catalytic residue His177 coordinated by Asn175 and Glu211 may orient and activate a water molecule to hydrolyze the enzyme-anhydride intermediate and release the free fatty acid (Figure 3.6C).

Acknowledgments

This work was supported by the U.S. National Science Foundation through its Engineering Research Center Program (Award No. EEC-0813570), leading to the Center for Biorenewable Chemicals (CBiRC), headquartered at Iowa State University and including Rice University, the University of California, Irvine, the University of New Mexico, the University of Virginia, and the University of Wisconsin-Madison. The authors thank M. Ann D.N. Perera and Zhihong Song of the W.M. Keck Metabolomics Research Laboratory at Iowa State University for assistance with fatty acid analysis, Paul Kapke and Amanda Brockman of the Hybridoma Facility at the Iowa State University for making the CvFatB2 antibody, Joel Nott of the Protein Facility at the Iowa State University for assistance with the CD spectrum. We thank Jarmila Tvaruzkova for generating CvFatB1 mutants. We also thank Colin Hueser for helping with the generation of CvFatB2 mutants and purification of CvFatB2 mutant proteins.

References

Cantu, D. C., Chen, Y. and Reilly, P. J. (2010). Thioesterases: a new perspective based on their primary and tertiary structures. Protein Science 19(7): 1281-1295.

Davies, H. M., Anderson, L., Fan, C. and Hawkins, D. J. (1991). Developmental Induction, Purification, and Further Characterization of 12-0-Acp Thioesterase

from Immature Cotyledons of Umbellularia-Californica. Archives of Biochemistry and Biophysics 290(1): 37-45.

Ginalski, K. (2006). Comparative modeling for protein structure prediction. Current Opinion in Structural Biology 16(2): 172-177.

Hawkins, D. J. and Kridl, J. C. (1998). Characterization of acyl-ACP thioesterases of mangosteen (Garcinia mangostana) seed and high levels of stearate production in transgenic canola. Plant Journal 13(6): 743-752.

Hedstrom, L. (2002). Serine protease mechanism and specificity. Chemical Reviews 102(12): 4501-4524.

Hills, M. J. (1999). Improving oil functionality by tuning catalysis of thioesterase. Trends in Plant Science 4(11): 421-422.

Jing, F., Cantu, D. C., Tvaruzkova, J., Chipman, J. P., Nikolau, B. J., Yandeau-Nelson, M. D. and Reilly, P. J. (2011). Phylogenetic and experimental characterization of an acyl-ACP thioesterase family reveals significant diversity in enzymatic specificity and activity. BMC Biochemistry 12: 44.

Jones, A., Davies, H. M. and Voelker, T. A. (1995). Palmitoyl-acyl carrier protein (ACP) thioesterase and the evolutionary origin of plant acyl-ACP thioesterases. Plant Cell 7(3): 359-371.

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. Journal of Molecular Biology 292(2): 195-202.

Lennen, R. M. and Pfleger, B. F. (2012). Engineering Escherichia coli to synthesize free fatty acids. Trends in Biotechnology 30(12): 659-667.

Mayer, K. M. and Shanklin, J. (2005). A structural model of the plant acyl-acyl carrier protein thioesterase FatB comprises two helix/4-stranded sheet domains, the N-terminal domain containing residues that affect specificity and the C-terminal domain containing catalytic residues. Journal of Biological Chemistry 280(5): 3621-3627.

Ohlrogge, J. B. (1994). Design of New Plant Products: Engineering of Fatty Acid Metabolism. Plant Physiology 104(3): 821-826.

Pidugu, L. S., Maity, K., Ramaswamy, K., Surolia, N. and Suguna, K. (2009). Analysis of proteins with the 'hot dog' fold: Prediction of function and identification of catalytic residues of hypothetical proteins. BMC Structural Biology 9.

Pollard, M. R., Anderson, L., Fan, C., Hawkins, D. J. and Davies, H. M. (1991). A Specific Acyl-Acp Thioesterase Implicated in Medium-Chain Fatty-Acid

Production in Immature Cotyledons of Umbellularia-Californica. Archives of Biochemistry and Biophysics 284(2): 306-312.

Song, F., Thoden, J. B., Zhuang, Z., Latham, J., Trujillo, M., Holden, H. M. and Dunaway-Mariano, D. (2012). The catalytic mechanism of the hotdog-fold enzyme superfamily 4-hydroxybenzoyl-CoA thioesterase from Arthrobacter sp. strain SU. Biochemistry 51(35): 7000-7016.

Thoden, J. B., Holden, H. M., Zhuang, Z. and Dunaway-Mariano, D. (2002). X-ray crystallographic analyses of inhibitor and substrate complexes of wild-type and mutant 4-hydroxybenzoyl-CoA thioesterase. Journal of Biological Chemistry 277(30): 27468-27476.

Thoden, J. B., Zhuang, Z., Dunaway-Mariano, D. and Holden, H. M. (2003). The structure of 4-hydroxybenzoyl-CoA thioesterase from arthrobacter sp. strain SU. Journal of Biological Chemistry 278(44): 43709-43716.

Voelker, T. (1996). Plant acyl-ACP thioesterases: chain-length determining enzymes in plant fatty acid biosynthesis. Genetic Engineering 18: 111-133.

Voelker, T. A., Hayes, T. R., Cranmer, A. M., Turner, J. C. and Davies, H. M. (1996). Genetic engineering of a quantitative trait: Metabolic and genetic parameters influencing the accumulation of laurate in rapeseed. Plant Journal 9(2): 229-241.

White, S. W., Zheng, J., Zhang, Y. M. and Rock (2005). The structural biology of type II fatty acid biosynthesis. Annual Review of Biochemistry 74: 791-831.

Yuan, L., Nelson, B. A. and Caryl, G. (1996). The catalytic cysteine and histidine in the plant acyl-acyl carrier protein thioesterases. Journal of Biological Chemistry 271(7): 3417-3419.

Zhang, X. J., Li, M., Agrawal, A. and San, K. Y. (2011). Efficient free fatty acid production in Escherichia coli using plant acyl-ACP thioesterases. Metabolic Engineering 13(6): 713-722.

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9: 40.

Zhuang, Z., Latham, J., Song, F., Zhang, W., Trujillo, M. and Dunaway-Mariano, D. (2012). Investigation of the catalytic mechanism of the hotdog-fold enzyme superfamily Pseudomonas sp. strain CBS3 4-hydroxybenzoyl-CoA thioesterase. Biochemistry 51(3): 786-794.

Figures

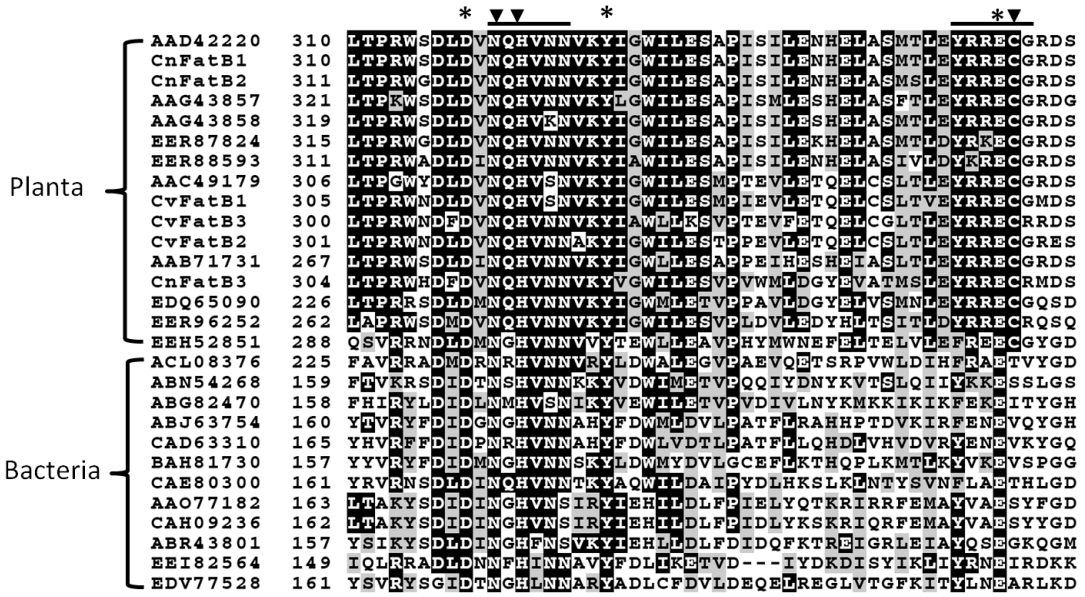


Figure 3.1 A 50-amino acid portion of a multiple sequence alignment of 28 acyl-ACP thioesterases. Two active-site motifs are indicated by lines. The previously proposed catalytic residues (N311, H313, and C348 relative to the sequence of CvFatB2) are indicated by arrowheads. Other conserved residues (D309, Y319, and E347) in this region are indicated by asterisks.

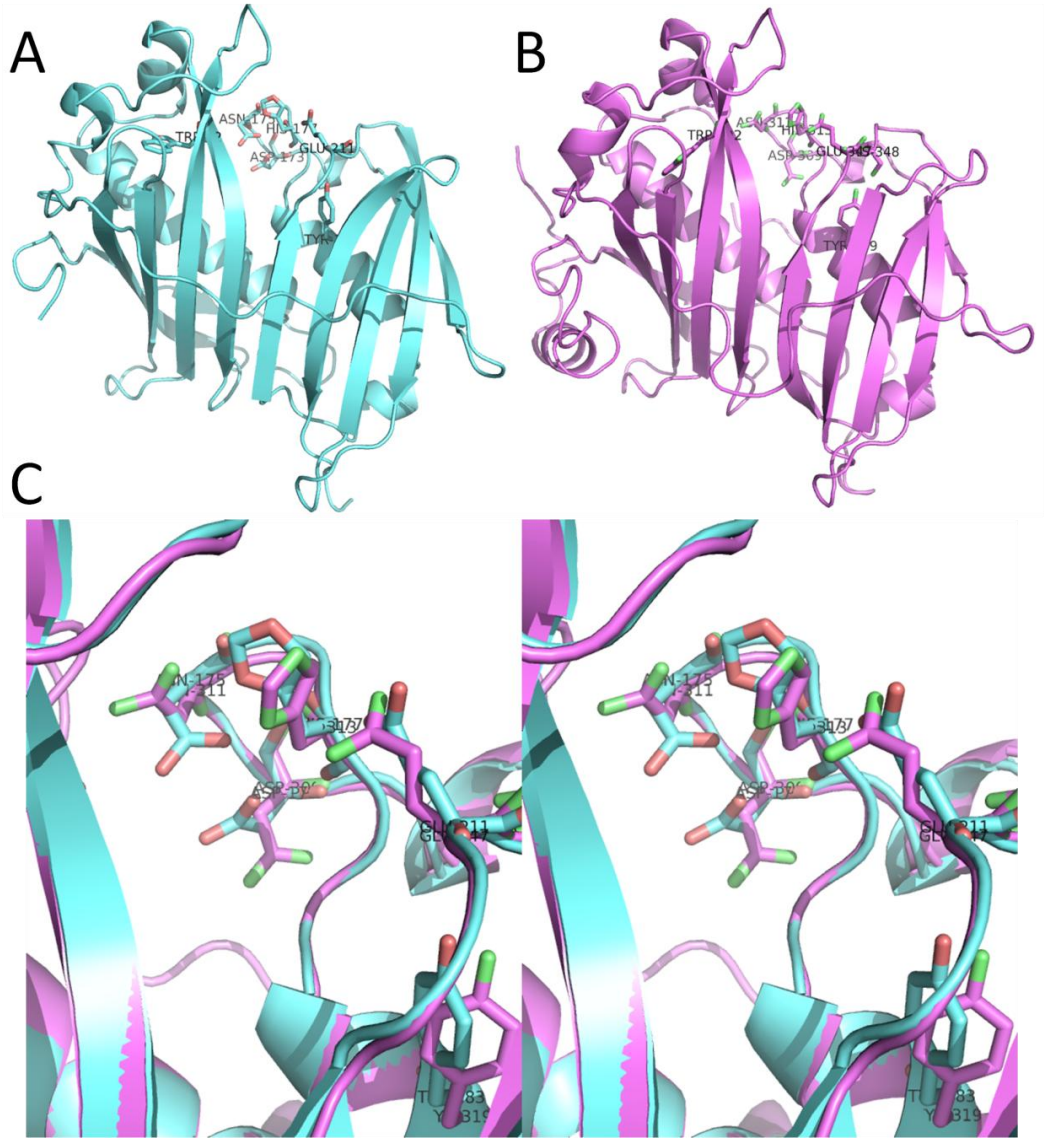


Figure 3.2 Crystal structure of a bacterial TE (2OWN) (A) and the predicted structural model for CvFatB2 (B), and a stereo view of the active sites of 2OWN and CvFatB2 (C). Trp192, N309, N311, H313, Y319, and E347 in CvFatB2 (shown as stick models) are the conserved residues identified from the multiple sequence alignment of 28 TEs. They have similar orientations as those corresponding residues in the L. plantarum acyl-ACP TE structure, 2OWN. Residues N311, H313, and C348 in CvFatB2 are previously proposed catalytic residues. C348 orients distantly from the other two catalytic residues in the CvFatB2 model and no Cysteine can be identified near the active sites in 2OWN structure.

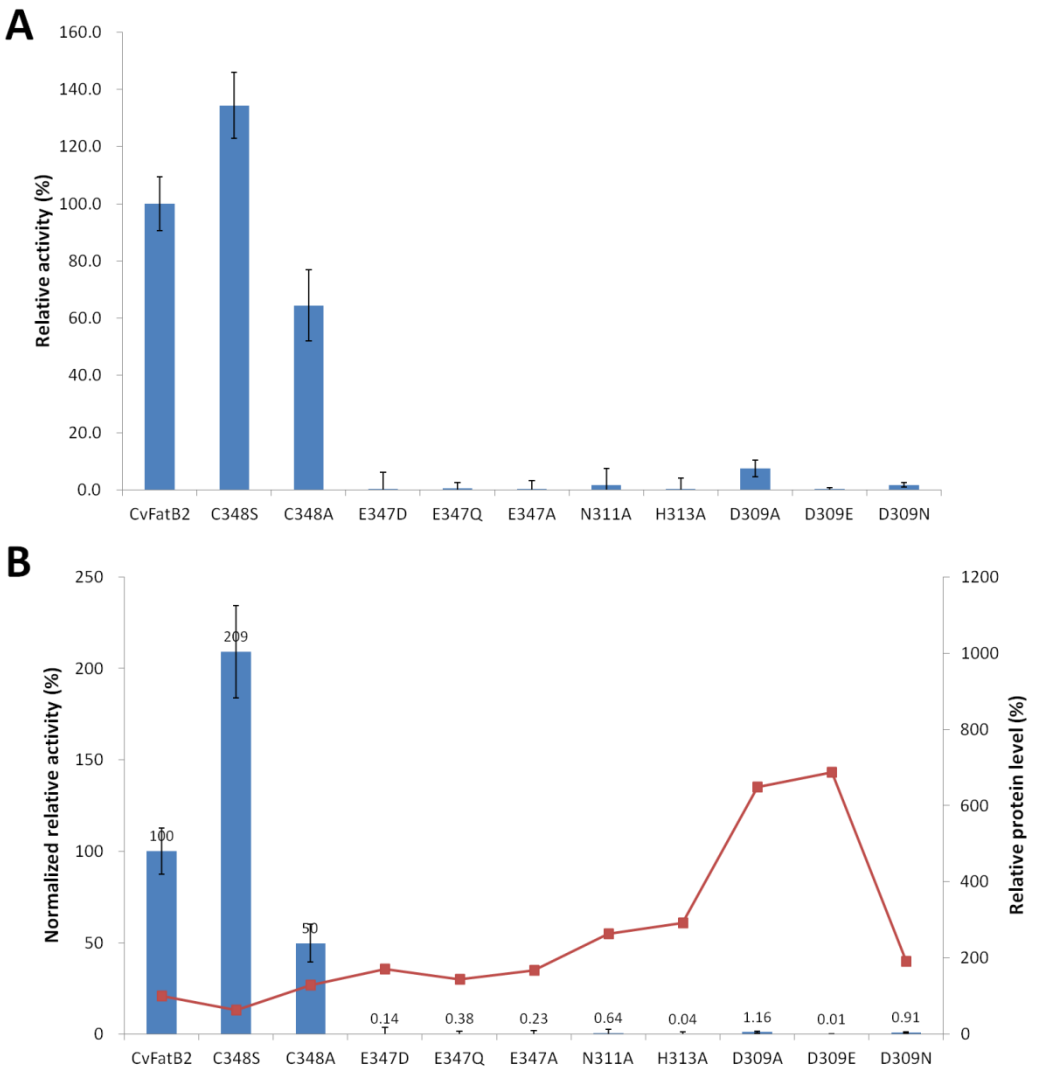


Figure 3.3 Relative activities (A) and normalized relative activities of CvFatB2 mutants (B). Each wild-type and mutant acyl-ACP TE was expressed in *E coli* strain K27. The accumulated free fatty acids were analyzed by GC-MS and used to represent the *in vivo* TE activity. The activity of each mutant was normalized to the activity of wild-type TE, resulting in relative activity. The protein expression levels of wild-type and mutant TEs were assessed via western analyses. The protein expression level of CvFatB2 in *E. coli* strain K27 was defined as 100% and the relative expression level of each mutant protein was calculated using the ratio of band intensity of mutants and wild-type. To determine the normalized relative activity, the total fatty acid accumulation of each mutant was first normalized to the protein expression level and then normalized to wild-type TE activity. Four replicates for each TE were analyzed.

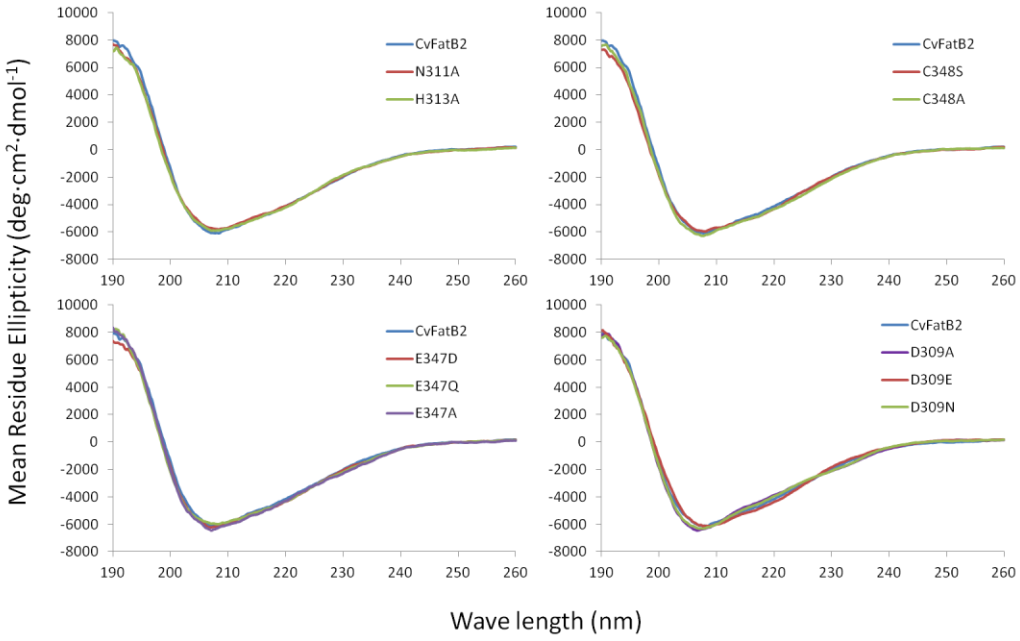


Figure 3.4 CD spectra of CvFatB2 and its mutants. The spectral shapes of site-directed mutants are very similar to the wild-type protein CvFatB2, suggesting the mutations did not change the folding.

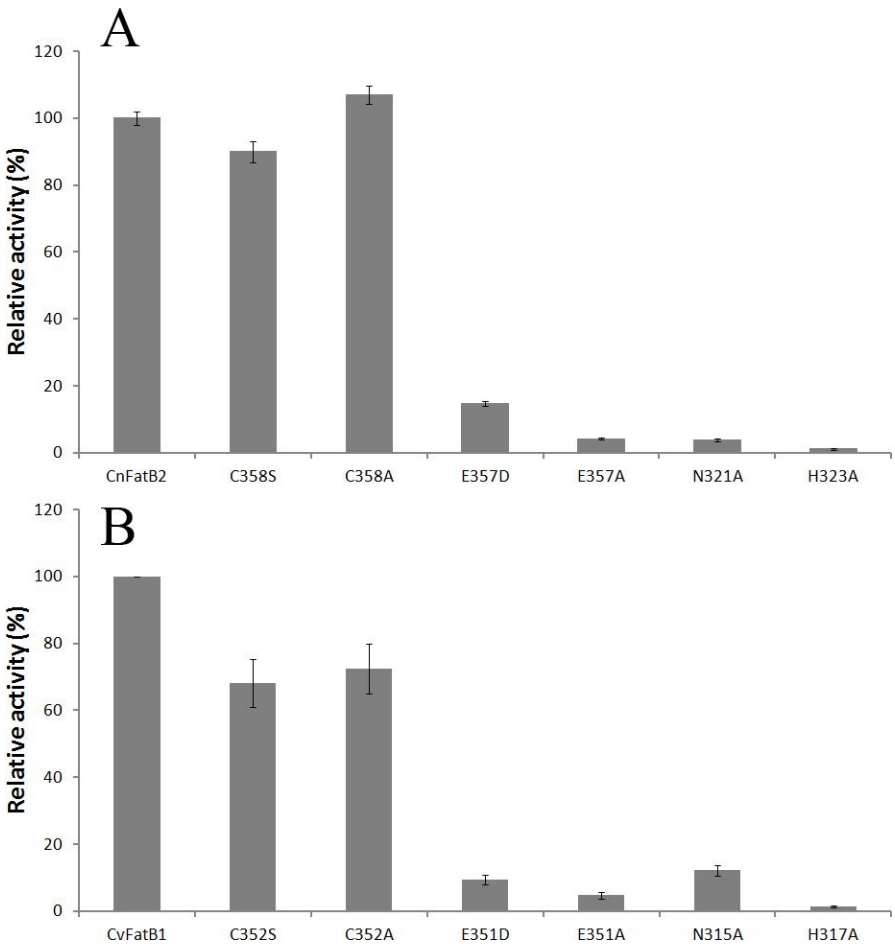


Figure 3.5 Relative activities of site directed mutants of CnFatB2 (A) and CvFatB1 (B). Each wild-type and mutant TE was expressed in *E coli* strain K27. The accumulated free fatty acids were analyzed by GC-MS and used to represent the *in vivo* TE activity. The activity of each mutant was normalized to the activity of the corresponding wild-type TE, resulting in relative activity. Four replicates for each TE were analyzed.

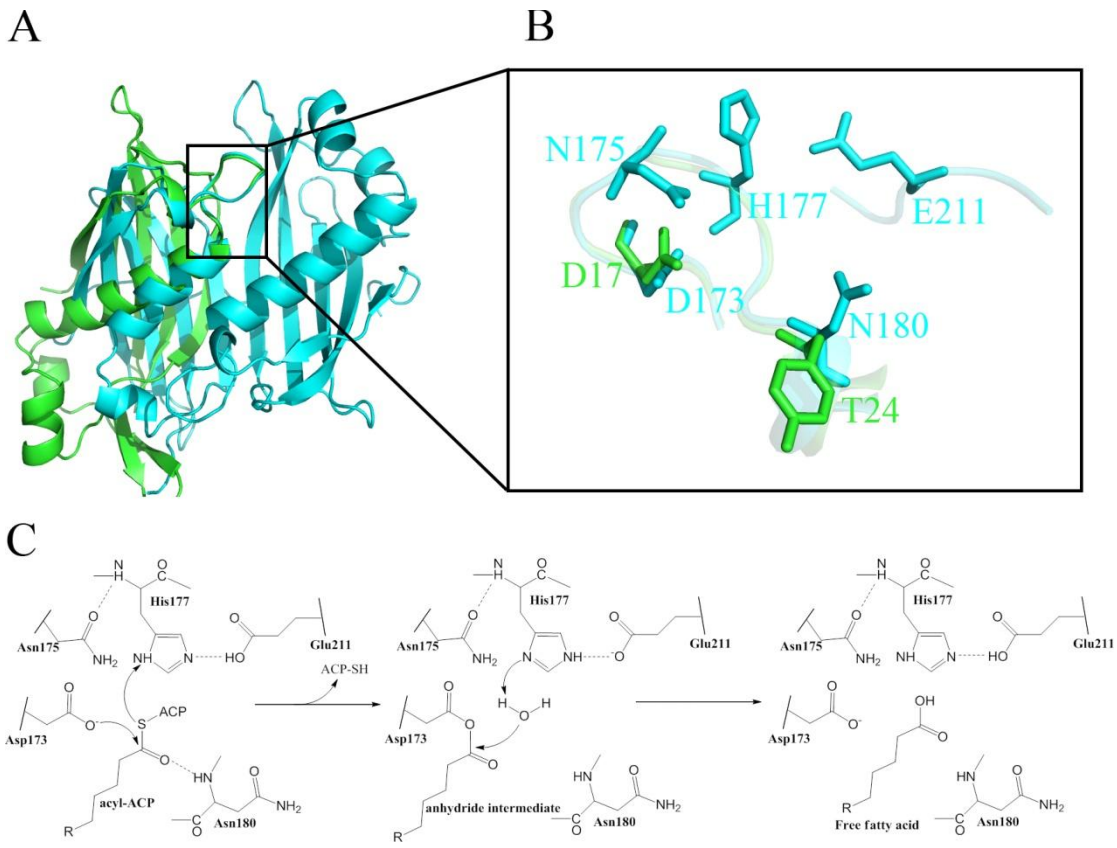


Figure 3.6 Structure comparison between *L. plantarum* acyl-ACP TE (2OWN) and *Pseudomonas* 4HBT (1BVQ) and the proposed catalytic mechanism for acyl-ACP TE. A, alignment of the active site structures of 2OWN (cyan) and 1BVQ (green); the loops structure in the box containing the catalytic residues align perfectly. B, catalytic residues of 2OWN (cyan) and 1BVQ (green) shown as stick models; C, proposed catalytic mechanism for acyl-ACP TE. Asp173, Asn175, His177, and Glu211 in 2OWN are equivalent residues of Asp309, Asn311, His313, and Glu347 in CvFatB2.

Supplemental data

Figure S3.1 A portion of a multiple sequence alignment of 28 acyl-ACP thioesterases. Residue Trp162 (relative to the sequence of CvFatB2) is conserved among all the sequences (asterisk).

Table S3.1 Secondary structures of CvFatB2 and mutants calculated from CD spectra data. Jfit was used to calculate the secondary structures. The *p*-value was obtained by χ^2 test.

Samples	Calculated secondary structures (%)			*p*-value compared with CvFatB2
	Helix	Sheet	Coil	
CvFatB2	11.9	24.0	64.1	
D309A	9.6	26.9	63.6	0.28
D309E	13.9	21.0	65.1	0.30
D309N	10.4	25.7	63.9	0.59
N311A	11.1	25.6	63.3	0.75
H313A	11.5	24.6	63.9	0.96
E347A	10.2	26.7	63.1	0.41
E347D	10.5	26.4	63.2	0.51
E347Q	9.7	27.8	62.5	0.18
C348A	10.3	26.8	62.9	0.40
C348S	11.1	25.6	63.3	0.77
PSIPRED model	11.9	28.5	59.6	0.15

CHAPTER 4 – DISSECTING THE STRUCTURAL DETERMINANTS OF CHAIN LENGTH SELECTIVITY OF PLANT ACYL-ACP THIOESTERASES

Fuyuan Jing, Marna D. Yandeau-Nelson, Basil J. Nikolau
Department of Biochemistry, Biophysics, and Molecular Biology, Iowa State University

A manuscript to be submitted for publication

Abstract

The substrate specificity of acyl-ACP TE plays a critical role in determining the chain length of fatty acids produced by the fatty acid synthesis (FAS) pathway. Understanding the structural basis of substrate specificity of acyl-ACP TE is beneficial for tailoring this enzyme with desired specificity for the application in metabolic engineering. To identify the region(s) that determine the substrate specificity, two acyl-ACP TEs (CvFatB1 and CvFatB2) from *Cuphea viscosissima*, having high amino acid sequence identity but different substrate specificities, were used for domain-shuffling study. In two rounds of domain shuffling, a total of 18 chimeric TEs were constructed and *in vivo* characterized in *E. coli* strain K27. Comparison of the substrate specificities of these chimeric TEs indicated that Fragment III is the most import region for determining the specificity, Fragments I, II and IV affect specificity to a less extent, and Fragments V and VI are not involved in specificity determination. Structural modeling of CvFatB2 revealed a hotdog structure. The N-terminal hotdog domain comprises mostly Fragments II and III, and is therefore the structural basis for the substrate specificity. Based on a multiple sequence alignment of 21 functionally characterized plant acyl-ACP TEs, nine residues were predicted to affect the substrate specificity. Together with the residues in Fragment III that vary in CvFatB1 and CvFatB2, 15 residues in the CvFatB2 sequence were subjected to site-directed mutagenesis analysis to verify their effects on substrate specificity. Comparison of the fatty acid profiles of those site-directed mutants demonstrated that six residues play critical roles in determining the substrate specificity, including V194 in Fragment II, V217, N223, R226, and R227 in

Fragment III, and I268 in Fragment IV. Another three residues, L257, I260, and L289, were proved to impact the catalytic activity of acyl-ACP TE, because they are in two proposed ACP binding motifs. Analysis of a crystal structure of acyl-ACP TE from *Lactobacillus plantarum* (2OWN) and comparison with the experimental data in this study lead to a hypothesis that plant acyl-ACP TE has a similar cavity in the N-terminal hotdog domain between the central α-helix and the anti-parallel β-sheets, which determines the substrate specificity.

Introduction

De novo fatty acid biosynthesis in higher plants is mainly located in the plastids (Rawsthorne 2002). This metabolic process, commonly primed with acetyl-CoA, involves reiterative condensation of the malonyl-moiety from acyl carrier protein (ACP), resulting in two-carbon unit extension of the acyl chain during each reaction cycle (Ohlrogge et al. 1997). The fatty acid elongation cycle is catalyzed by type II fatty acid synthase (FAS II), which integrates the capabilities of discrete enzymes that catalyze each reaction in the elongation cycle (White et al. 2005). Acyl-ACP thioesterase (TE) catalyzes the terminal reaction in the plastidial fatty acid biosynthesis pathway, which is the hydrolysis of the thioester bond of acyl-ACPs to release free fatty acids and ACP. Fatty acids are then transported into the cytosol, esterified to coenzyme A, and further incorporated into phospholipids, triacylglycerols, and other neutral lipids. Acyl-ACP TE plays a crucial role in determining the chain length of the fatty acids produced by the FAS system, and thus the fatty acid composition within membrane and storage lipids.

Plant acyl-ACP TEs are encoded by nuclear genes, and contain ~60 amino acid transit peptides that target plastid localization (Voelker et al. 1992). Based on protein sequence alignment as well as known substrate preferences, acyl-ACP TEs are classified into two groups, FatA and FatB (Jones et al. 1995). FatA TEs specifically hydrolyze the unsaturated fatty acid substrate, 18:1-ACP. In contrast, FatB TEs preferentially act on saturated acyl-ACPs with acyl chains varying from 8- to 18-carbons in length. A prior functional characterization of 31 acyl-ACP TEs from both plants and bacteria revealed

that the FatB-type TEs can be further clustered into three classes based on their substrate specificities: Class I TEs that mainly act on 14- and 16-carbon acyl-ACPs, Class II TEs that have broad substrate specificities, and Class III TEs that act predominantly on 8-carbon acyl-ACPs (Jing et al. 2011). It has been suggested that the TE that acts on 16:0-carbon acyl-ACP is ancient and ubiquitous in plants (Jones et al. 1995). Other specialized TEs that act on medium-chain acyl-ACPs have evolved from these enzymes and are responsible for the accumulation of medium chain fatty acids in the seeds of some species, including *Umbellularia californica*, *Cuphea hookeriana*, and *Cocos nucifera* (Davies et al. 1991; Voelker et al. 1992; Dehesh et al. 1996).

Acyl-ACP TEs are of biotechnological interest because of their application in the genetic engineering of oil crops to modify the fatty acid composition of the seed oil for improved nutritional or industrial value (Thelen et al. 2002). For example, the acyl-ACP TE, UcFatB1 from *U. californica* was overexpressed in oilseed rape, leading to significant accumulation (over 50%) of 12:0 fatty acid in seed oil (Voelker et al. 1996), which has many industrial applications but is not commonly produced by US-based oil crops. Similarly, overexpression of the acyl-ACP TE, ChFatB2 from *C. hookeriana* induced high levels of accumulation of 8:0 and 10:0 fatty acids in seed oil of canola (Dehesh et al. 1996). Acyl-ACP TEs have recently gained more attention relative to the metabolic engineering of the FAS pathway in microbes or microalgae for the production of biorenewable chemicals and biofuels (Steen et al. 2010; Liu et al. 2011; Zhang et al. 2011; Lennen et al. 2012). For example, fatty acids of different chain lengths and substituted fatty acids (e.g. unsaturated and hydroxyl fatty acids) have diverse applications in the chemical industry (Nikolau et al. 2008). To produce each targeted fatty acid with high purity during fermentation, it is highly desired to have specific acyl-ACP TEs that have the capability to selectively hydrolyze the intermediates of the FAS pathway. The demand for specific acyl-ACP TEs has promoted the interest in understanding the fundamental mechanism that determines the substrate specificities of this class of enzymes.

Due to a lack of crystal structures for plant acyl-ACP TEs, previous studies to identify amino acids involved in substrate specificity mainly relied on domain shuffling between different acyl-ACP TEs and analyzing the substrate specificity of the resulting chimeric enzymes. By comparing two chimeric enzymes generated from a 12:0-ACP TE (UcFatB1) from *U. californica* and a 14:0-ACP TE (CcFatB1) from *Cinnamomum camphorum*, it was concluded that the C-terminal two-thirds of the protein determines substrate specificity (Yuan et al. 1995). Moreover, a double mutant (M197R-R199H) within this region transformed the substrate specificity of UcFatB1 from 12:0-ACP to 12:0/14:0-ACP with equal preference for both substrates (Yuan et al. 1995). In contrast, characterization of chimeric enzymes generated from the Arabidopsis acyl-ACP TEs AtFatA and AtFatB demonstrated that the N-terminus and not the C-terminus of these proteins is critical for determination of substrate specificity (Salas et al. 2002). Structural modeling of Arabidopsis AtFatB revealed that this plant acyl-ACP TE adopts a hotdog structure and the N-terminal domain contains residues that affect substrate specificity (Mayer et al. 2005; Mayer et al. 2007). Although some success has been achieved in identifying amino acid residues important for substrate specificity, the complete structural basis for specificity remains obscure.

As more acyl-ACP TEs have been functionally characterized, there is a greater opportunity to utilize the power of bioinformatic analysis on these diverse TE sequences and diverse specificities to identify amino acids that determine the substrate specificity. Two acyl-ACP TEs from *Cuphea viscosissima* (CvFatB1and CvFatB2) share more than 70% identity in amino acid sequences. When expressed in *E. coli* strain K27, CvFatB1 produced mainly 8- and 10-carbon fatty acids, while CvFatB2 produced primarily 14- and 16-carbon fatty acids (Jing et al. 2011). In this study, two complementary approaches were taken to identify residues that govern this difference in substrate specificity between the two acyl-ACP TEs. First, domain shuffling within the CvFatB1 and CvFatB2 acyl-ACP TEs was used to determine the region(s) of the enzyme that affect the substrate specificity. Second, nine specific residues that may affect substrate specificity were identified from sequence alignment of functionally characterized acyl-

ACP TEs and structural modeling of CvFatB2. Site-directed mutagenesis was used to test whether these predicted residues determine the substrate specificity of these enzymes. Six residues proved to affect the substrate the specificity of acyl-ACP TE.

Methods

Construction of chimeric acyl-ACP TEs

Previously, the mature portions of the CvFatB1 and CvFatB2 cDNAs (i.e. without plastid transit peptides) were codon optimized, synthesized, and cloned into pUC57 vector under the control of the lacZ promoter. Using the primers listed in table 4.1, six fragments (I, II, III, IV, V, and VI) for each TE gene were generated. To construct the chimeric acyl-ACP TEs, six fragments either from CvFatB1 or CvFatB2 were assembled by overlap extension PCR (Figure 4.1). PCR was first performed in a 50-μL reaction mixture containing 10 ng of each fragment, 1x Phusion buffer, 0.2 mM dNTP, and 1 Unit of Phusion high-fidelity DNA polymerase (New England Biolabs, USA) using a cycling program of 98 ℃ for 2 min, 8 cycles of 98 ℃ for 10 s, 50 ℃ for 15 s and 72 ℃ for 20 s, and a final extension step of 72 ℃ for 5 min. Upon completion of this PCR program, 0.5 uM pUC57F and pUC57R primers were immediately added into the reaction mixture, followed by a second PCR cycling program of 98 ℃ for 2 min, 30 cycles of 98 ℃ for 10 s, 54 ℃ for 15 s and 72 ℃ for 20 s, and a final extension step of 72 ℃ for 5 min. The targeted full-length gene products were separated by electrophoresis on a 1% agarose gel, recovered from the gel using the QiaQuick gel extraction kit (Qiagen, Valencia, CA, USA) and cloned into the pUC57 vector using the BamHI and EcoRI restriction sites. The sequence of each construct was confirmed by sequencing with primers pUC57F and pUC57R.

Multiple sequence alignment of functionally characterized acyl-ACP TEs

A total of 21 functionally characterized plant acyl-ACP TEs were collected and separated into two groups based on their substrate specificities. Group A includes 14

acyl-ACP TEs that showed substrate specificity towards C14/16-ACPs and were hypothesized to be the ancient and ubiquitous enzymes (Jones et al. 1995). Group B contains 7 acyl-ACP TEs that had activity on 8- to 12-carbon acyl-ACPs and are thought to have evolved from group A acyl-ACP TEs. The protein sequences of these acyl-ACP TEs were subjected to multiple sequence alignment using Vector NTI (Invitrogen, USA) with default parameters.

Structural modeling of CvFatB2

The mature portions of the CvFatB2 protein sequences were subjected to tertiary structure calculation with I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/), using a bacterial crystal structure of acyl-ACP TE from *Lactobacillus plantarum* (2OWN) as a template (Zhang 2008). The resulting tertiary structures were viewed and analyzed with the PyMOL Molecular Graphics System, Version 1.5.0.4 (Schrödinger, LLC).

Site-directed mutagenesis

To test whether the predicted residues affect the substrate specificity of acyl-ACP TEs, point mutations were generated in CvFatB2 using the QuikChange II site-directed mutagenesis kit (Agilent Technologies, USA), according to manufacturer instructions. Multiple mutations were generated sequentially. The targeted residues in CvFatB2 were mutated to the corresponding residues in CvFatB1. The authenticity of all site-directed mutants was confirmed by sequencing both strands of the plasmids.

In vivo activity of acyl-ACP TE variants

Chimeric acyl-ACP TEs and site-directed mutants were expressed in *E coli* strain K27, which contains a mutation in the fadD gene that disrupts beta-oxidation and leads to accumulation of free fatty acids. Free fatty acids that accumulated in the medium were extracted and analyzed (Jing et al. 2011). For each TE construct, four individual colonies

were cultured in 2 mL LB medium supplemented with 100 mg/L carbenicillin. When the culture reached an OD600 of ~0.7, the growth medium was replaced with 3 mL of M9 minimal medium (47.7 mM Na2HPO4, 22.1 mM KH2PO4, 8.6 mM NaCl, 18.7 mM NH4Cl, 2 mM MgSO4, and 0.1 mM CaCl2) supplemented with 0.4% glucose and 100 mg/l carbenicillin. Acyl-ACP TE expression was induced by adding 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG). After 40 hours cultivation at 30 ℃, free fatty acids were extracted from the medium and analyzed via GC-MS. Fatty acid production for each acyl-ACP TE was measured by subtracting the fatty acids produced by *E. coli* expressing a control plasmid (pUC57) that lacks the TE gene. The mol percentage of individual fatty acids was calculated as a reflection of the substrate specificity of expressed acyl-ACP TE. Fatty acid production of an acyl-ACP TE is the total amount (μM) of free fatty acids produced, and fatty acid profile is the mol percentage of each individual fatty acid.

Statistical cluster analysis

To classify acyl-ACP TE variants based on their *in vivo* activities, the fatty acid profile data obtained from the *in vivo* expression of all TE variants were used to perform statistical clustering analysis. The distance matrix was calculated using Euclidean distances, and Ward's method (Ward 1963) was used to perform agglomerative hierarchical clustering. The *p*-values were calculated via multiscale bootstrap resampling with 1000 replicates (Suzuki et al. 2006).

Determination of acyl-ACP TE protein expression level

To examine the protein expression levels of acyl-ACP TE variants, soluble proteins were extracted from cell pellets and analyzed by immune-blotting (Chapter 3). For each acyl-ACP TE, 35 μg crude extract was separated on SDS-PAGE gel, transferred onto nitrocellulose membrane, and analyzed via western blot using anti-CvFatB2 antibody.

Results

Substrate specificity of chimeric acyl-ACP TEs

CvFatB1 and CvFatB2 share more than 70% identity in amino acid sequences (Figure 4.1) but exhibit different substrate specificities when expressed in *E. coli* strain K27. We hypothesize that a subset of the ~80 different residues between these two proteins may determine their different substrate specificities. To identify the region or residues that impact substrate specificity, CvFatB1 and CvFatB2 were separated into 6 fragments by PCR amplification and used to generate chimeric acyl-ACP TEs. It would need 62 (2^6-2) chimeras to cover all the possible recombination of 6 fragments in two different acyl-ACP TEs. To narrow the targeted region using a minimal number of chimeras, we did two rounds of domain swapping. For each chimeric acyl-ACP TE, a six-digit binary number was used to represent the six fragments I to VI. Number"0" indicates a fragment from CvFatB1, while number "1" indicates a fragment from CvFatB2. The six-digit binary numbers were converted to decimal numbers to name the chimeric TEs (Figure 4.2).

In the first round of domain shuffling, six chimeric acyl-ACP TEs, including rTE3, rTE12, rTE15, rTE48, rTE51, and rTE60, were generated by domain swapping of three combined fragments, I-II, III-IV, and V-VI (Figure 4.3). Each chimeric acyl-ACP TE, as well as wild-type CvFatB1 and CvFatB2, was expressed in *E. coli* strain K27 and fatty acid products were analyzed. The composition of fatty acid product for each TE was calculated as a representation of their substrate specificities (Figure 4.3). Chimeric rTE3, which resulted from replacing fragment V-VI in CvFatB1, displayed almost the same substrate specificity as CvFatB1. Similarly, replacement of fragment V-VI in CvFatB2 resulted in chimeric rTE60, which showed the same substrate specificity as CvFatB2. In contrast, the other chimeric TEs with substitution of fragment I to VI (rTE12, rTE15, rTE48, and rTE51) exhibited substrate specificities between that of CvFatB1 and CvFatB2 (Figure 4.3). The comparison of substrate specificities of these

chimeric and wild-type TEs suggested that the residues that determine substrate specificity reside within fragments I, II, III and IV.

To further narrow the region that determines substrate specificity, a second round of domain swapping was carried out, using only fragments I to IV. Another 12 chimeric TEs were generated and characterized in *E. coli* strain K27. Their substrate specificities are shown in Figure 4.4. The chimeric constructs, rTE8, rTE24, rTE28, rTE40, rTE44, rTE56, and rTE60 each contain fragment III from CvFatB2 and like CvFatB2, mainly produce 14- to 16-carbon fatty acids (i.e., more than 50% of these fatty acids) (Figure 4.4). In contrast, chimeric constructs, rTE4, rTE16, rTE20, rTE32, rTE36, and rTE52 each contain fragment III from CvFatB1, and these proteins produce 8- to 12-carbon fatty acids (like CvFatB1) (Figure 4.4). Chimeric rTE8 differs from CvFatB1 only within fragment III, which contains CvFatB2 sequence and exhibits dramatically different fatty acid composition as compared to CvFatB1. By replacing only fragment III in rTE60 with CvFatB1 sequence, the chimeric rTE52 showed a significant change in substrate specificity from C14/16 to C8/10/12 specificity. Together, these results indicate that fragment III contains the most important residues that determine the substrate specificity of acyl-ACP TE. Comparison of the substrate specificities of other chimeric TEs suggests that fragments I, II, and IV also affect the substrate specificity, but to a lesser extent. A comparison of CvFatB1 and CvFatB2 primary sequence reveals only ten residues that differ between the two TEs within fragment III, which can be individually characterized via site-directed mutagenesis.

Predicted tertiary structures of CvFatB2

Although several acyl-ACP TEs from bacteria have known crystal structures, none exist for plant sourced enzymes. Therefore, we predicted the tertiary structure of CvFatB2 (Figure 4.5) using a *L. plantarum* acyl-ACP TE structure (2OWN) in an effort to understand the structural basis for the substrate specificity. The predicted CvFatB2 structure contains two hotdog domains linked by a long coil and is similar to a previously predicted structure model for Arabidopsis acyl-ACP TE (Mayer et al. 2005).

When mapping the six fragments of acyl-ACP TE to the predicted structure, we found that the majority of the N-terminal hotdog structure consists of fragments II and III, and the C-terminal hotdog structure consists of fragments V and VI (Figure 4.5). Fragment I forms mostly the coil structure at the very N-terminus and one β-sheet structure in the hotdog domain. Fragment IV contains a small α-helix in the N-terminal domain, a β-sheet in the C-terminal domain, and mostly a long coil structure that spans across the two hotdog domains. Fragment III, which this study establishes as being key to determining the substrate specificity of acyl-ACP TE, forms three β-sheets that wrap the central α-helix. Based on the observation of a cavity between the β-sheets and α-helix in the 2OWN crystal structure from *L. plantarum*, we further propose that plant acyl-ACP TE may also form a cavity in the N-terminal hotdog structure, though the predicted CvFatB2 structure may not be sufficiently accurate to visualize that structure.

Sequence alignment of functionally characterized acyl-ACP TEs

To further pinpoint the possible residues that may affect the substrate specificity of acyl-ACP TEs, 14 plant acyl-ACP TEs in group A that display preference toward C14/16-ACPs and 7 TEs in group B that are specific for C8- to C12-ACPs were subjected to multiple sequence alignment (Figure S4.1). Based on our finding that fragments I to IV contain residues determining the substrate specificity, only the residues within Fragments I-IV of the alignment were examined for possible critical residues that determine substrate specificity. A residue is predicted as a candidate according to the following criteria: (1) it is conserved in at least one TE group (ie. group A or B), which indicates an important role in the functionality of that group; (2) it is different between the two groups in certain amino acid properties (i.e. size, charge, hydrophobicity), which suggests it may contribute to the difference in substrate specificity between the groups. Using this strategy, 9 residues were predicted to affect substrate specificity: V194, V217, N223, L257, I260, I268, A276, D282, and L289 (numbered based on the sequence of CvFatB2 (GenBank AEM72523)) (Figure 4.6). Among these 9 predicted positions, residue 194 resides in fragment II, residues 217 and 223 are in fragment III, and the

remaining six residues are in fragment IV. At positions 194, 223, 257, and 268, Group A acyl-ACP TEs have conserved Val, Asn, Leu, and Ile, respectively, while in Group B enzymes these positions were occasionally substituted with Leu, Ile, Ile/Phe, and Leu/Phe, respectively (Figure 4.6). At position 217, Group B enzymes utilize larger amino acids (i.e. Ile, Leu, Phe) than those in Group A enzymes (i.e. Val and Ala). At position 289, Group B TEs consistently has a Phe residue, while group A TEs have a conserved Leu. Residues 260, 276, and 282 were selected because group B acyl-ACP TEs have relatively conserved amino acids at those positions, while in Group A enzymes these positions utilize more diverse set of residues.

Site-directed mutagenesis analysis

To experimentally test the functionality of the identified critical residues in Fragments II, III and IV that may determine substrate specificity of acyl-ACP TEs, site-directed mutagenesis was performed on the 15 residues that vary between CvFatB1 and CvFatB2 within this region. One of these tested residues resides in Fragment II, 8 residues reside in Fragment III, and 6 residues reside in Fragment IV. A total of 46 mutants were generated in the CvFatB2 sequence, by substituting the corresponding amino acids from CvFatB1 at single or multiple positions simultaneously. Based on the fatty acid profiles (Figure S4.2) generated by *E. coli* cultures expressing these TE mutants, the amino acid substitution mutants could be separated into two classes of substrate specificities using cluster analysis (Figures 4.7). One group, which includes the parental CvFatB2 sequence and 14 mutants, show production of predominantly C14/16 fatty acids, characteristic of CvFatB2. Therefore one can conclude that the residues mutated in these 14 TEs are not significant in determining the acyl-chain length specificity of the CvFatB1, which is for C8/C10 acyl-ACPs. The second group of TEs contains CvFatB1, and this group of TEs produces significantly increased levels of 4- to 12-carbon fatty acids, and reduced amounts of C14/16 fatty acids (Figure 4.7 and S4.2).

Mutation of L257, I260, A276, D282, and L289 alone or in combination resulted in little change in the fatty acid profiles, as compared to the wild-type CvFatB2,

suggesting that these residues do not affect substrate specificity. Single mutation N223I from CvFatB2 increased production of 8:0 fatty acid from 5.4% to 17.2% of total fatty acids (Figure S4.2). The mutants (CvB2MT13, CvB2MT21, and CvB2MT28) that combined the N223I mutation with other mutations that do not affect substrate specificity on their own consistently show increased activity against 8:0-ACP, to 13%-17%. These results suggest that residue N223, which is in Fragment III, can affect substrate specificity to some degree. All the mutant TEs that contain mutation V217F show a dramatic decrease in the percentage of 14- to 16-carbon fatty acids and a concomitant increase in the percentage of 4- to 12-carbon fatty acids (Figure 4.7), implying residue V217 has significant impact on substrate specificity. Similarly, V174 and I286 are verified to affect substrate specificity significantly by comparing the fatty acid profiles of those mutants. Compared to CvB2MT10, CvB2MT20 which has two more mutations at positions R226 and R227 with substitution of Ala and Ser respectively showed increased 8:0 and 10:0 at the expense of 14:0 and 14:1. Same changes were also observed in mutant CvB2MT45 which has an additional mutation R226G compared to CvB2MT10. These results indicate that R226 and R227 are also important residues that determine the substrate specificity. The addition of two more mutations, A218S and K219Q, converting CvB2MT10 to CvB2MT19, and CvB2MT20 to CvB2MT30 did not change the resulting fatty acid profiles, indicating that residue A218 and K219 do not affect the substrate specificity. Using the same approach to compare other mutants, we find that residue V213 and V248 are not important in determining substrate specificity. Most significantly, 11 mutations can convert the C14/16-specific CvFatB2 to a C8/C10 specific enzyme (i.e. CvB2MT40) that produces up to 61.4% 8:0 and 12.5% 10:0. Indeed, CvB2MT40 produced even more 8:0 fatty acid than CvFatB1, perhaps because residue I268 was mutated to a residue of larger size (Phe) instead of the Leu that is present in CvFatB1. In summary, by comparing the fatty acid profiles of those site-directed mutants, six residues have been shown to play critical roles in determining of substrate specificity, including V194 in Fragment II, V217, N223, R226, and R227 in Fragment III, and I268 in Fragment IV. Other residues analyzed, V213, A218, K219,

V248, L257, I260, A276, D282, and L289, do not have significant impact on this characteristic.

The multiple sequence alignment of 21 functionally characterized plant acyl-ACP TEs, indicate that the Phe residue at position 289 (numbered in CvFatB2 sequence), is characteristic of group B TEs, whereas group A TEs have a conserved Leu at this position (Figure 4.6). Because Group A and B TEs express different substrate specificities, we originally hypothesized that this residue might impact substrate specificity. However, data presented herein establish that a L289F mutation in CvFatB2 does not impact substrate specificity. This highly conserved position in the two Groups of FatB TEs may play another important role in catalysis, determining catalytic activity. Indeed, the L289F mutation in CvFatB2 appears to increases the thioesterase activity as judged by the *in vivo* production of fatty acids, when this mutant enzyme was expressed in *E. coli.* Because mutation L289F is combined with other two mutations (A276P and D282N) in most of the mutants, we first compare the mutants having these additional triple mutations to the mutants without them (Table 4.2). Results show that the triple mutations can significantly increase the total fatty acid production (p=0.002 in paired-sample t test). Then, we compare the single mutation at each three positions to the wild-type CvFatB2. Mutations A276P (CvB2MT7) and D282N (CvB2MT8) both decrease the total fatty acid production. In contrast, single mutation L289F (CvB2MT9) could increase the fatty acid production dramatically. Based on all these results, we conclude that residue L289 may be related to the activity of acyl-ACP TE. Similar strategies were used to study the effect of other residues on the fatty acid production, resulting in the identification of two mutations, L257F and I260L, which also increases the *in vivo* fatty acid production, maybe an indication of increased acyl-ACP TE activity.

The changes in fatty acid production and profile observed across the TE point mutants might be attributed to either altered catalytic properties or modified expression level of these mutants. Protein expression levels of CvFatB2 and 7 representative mutants, which displayed diverse fatty acid production and profile, were analyzed via western blotting (Figure 4.8). CvFatB2, rTE15, rTE28, rTE52, CvB2MT9, CvB2MT20,

CvB2MT30, and CvB2MT38 produced 175.2, 56.3, 279.8, 874.3, 410.8, 691.2, 1030.5, and 1008.9 μM fatty acids respectively, and showed different fatty acid profiles when expressed in *E. coli* strain K27 (Figure 4.8). Because all TE genes in pUC57 vector were driven by a weak lacZ promoter, the protein could not be observed with regular staining approach, suggesting that the TEs are heterologously expressed at very low levels in *E. coli* strain K27 cells. Western blot analysis shows that CvFatB2, rTE52, CvB2MT9, CvB2MT20, CvB2MT30 have similar protein levels, although they have different fatty acid production and profiles (Figure 4.8). Compared to CvFatB2, rTE28 and CvMT38 showed higher fatty acid production, even though these constructs had lower TE protein levels (Figure 4.8). In contrast, rTE15 was expressed more highly but fatty acid production was lower than CvFatB2 Together, the data suggest that at such low protein expression level, the total fatty acid production is mainly affected by the activity of acyl-ACP TE and the major determinant for fatty acid profile of *E. coli* expressing a TE is the substrate specificity of that TE.

Insight into the structure of a bacterial acyl-ACP TE

The lack of crystal structures for any plant acyl-ACP TE hampers the understanding of the structural basis for the substrate specificity of TE. However, we were able to gain some insights into plant acyl-ACP TEs by comparing these enzymes to the crystal structure (2OWN) of a closely related bacterial acyl-ACP TE from *L. plantarum*. In Chapter 3, by analyzing this bacterial acyl-ACP TE structure, we were able to identify the catalytic residues for acyl-ACP TE. In this study, the tertiary structure of a plant acyl-ACP TE (CvFatB2) was predicted, using 2OWN as a template. Comparison of the substrate specificities of chimeric TEs generated from CvFatB1 and CvFatB2 demonstrates that the N-terminal domain determines the substrate specificity of acyl-ACP TE. A cavity in the N-terminal hotdog domain can be observed in the 2OWN structure (Figure 4.9), leading us to hypothesize that such a cavity in plant acyl-ACP TEs resides between the central α-helix and the anti-parallel β-sheets and may accommodate the acyl chain of substrates. Because the predicted CvFatB2 structure does

not allow us to observe this cavity, we used the structure of 2OWN to identify the cavity forming residues and mapped these to the CvFatB2 sequence via sequence alignment.

In the crystal structure 2OWN, residues I34, V38, S41, E42, S45, L50, V55, W62, V64, T65, G87, S88, A89, Y90, A95, Y96, R97, W116, L133, V134, Y137, and S139 line a cavity in the N-terminal hotdog domain that opens toward the catalytic residues (D173, N175, H177, N180, and E211) of the enzyme (Figure 4.9). Four residues, W62, R97, W116, and Y137, comprise more than 50% of the cavity surface area (Figure 4.9), suggesting that they play critical roles in defining the shape and size of the cavity. From the sequence alignment of 2OWN and CvFatB2, the corresponding residues in CvFatB2 are identified as W192, R227, W247 and I268. In the multiple sequence alignment of 21 plant acyl-ACP TEs (Figure S4.1), position 192 is 100% conserved as Trp, indicating it has an important role in the functionality of acyl-ACP TE. While at position 247, 19 sequences have a Trp and 2 sequences have substitution with another bulky residue, Tyr. Residue R227 is only conserved in group A TEs, which are specific for C14/16 acyl-ACP substrates, and position 268 is always shows a hydrophobic residue (e.g., Ile, Val, and Phe). Our site-directed mutagenesis experiments indicate that mutations of residue R227 and I268 changes the substrate specificity of CvFatB2 dramatically. This result suggests that these two residues (R227 and I268), like the corresponding residues (R97 and Y137) in 2OWN, may be important for the shape and size of the cavity in plant acyl-ACP TE.

In the structure 2OWN, another region that contributes to the shape of the cavity are the residues from 87 to 97, which form a small portion of two β-sheets connected by a turn structure. Within this region, 7 residues (G87, S88, A89, Y90, A95, Y96, and R97) are identified to reside near the bottom of the cavity (Figure 4.9), suggesting they may determine the size of the cavity. This region aligns with residues 217 to 227 of the CvFatB2 sequence, within fragment III, which we have shown is most important in determining the substrate specificity of the plant acyl-ACP TE. Site-directed mutagenesis within this region also demonstrates that four residues (V217, N223, R226, and R227) affect the substrate specificity significantly. Especially, mutating V217 to a

larger size residue Phe, which probably decreases the size of the cavity, could shift the substrate specificity to shorter chain length.

In the tertiary structure 2OWN, residue V64 is immediately adjacent to W116, which occupies large surface area of the cavity. Based on this observation, a hypothesis was proposed that replacement of residue V64 with a larger size residue (e.g., Leu, Ile, and Phe) would force W116 to shrink the cavity and shift the substrate specificity to a shorter chain length. We tested this hypothesis with CvFatB2 by mutating the corresponding residue V194 to Phe. As a result, the single mutation, V194F almost inactivated the enzyme. However, when combined with other mutations, mutation V194F can change the substrate specificity to 4:0, 6:0, and 8:0 with high activity. For example, CvB2MT40, which contains 11 point mutations, can produce 61.4% 8:0 and 12.5% 10:0 fatty acids (Figure 4.7 and S4.2). Replacement of the residue (Val) at position 194 with residue Phe resulted in CvB2MT38, which can produce 1008.9 uM total fatty acids with 37.2% 8:0 and 31.1% 6:0 (Figure 4.7 and S4.2). This is the highest activity for 6:0-ACP ever reported in plant acyl-ACP TEs.

In the structure 2OWN, the other five residues (I34, V38, S41, E42, and S45) in the central α-helix are also found around the opening of the cavity. The corresponding region in CvFatB2 is from 158 to 169, which belongs to fragment II. The domain-shuffling experiment has shown that fragment II also affects the substrate specificity, though not as much as fragment III. We hypothesize that residues between positions 158 and 169 may be involved in the determination of substrate specificity. However, site-directed mutagenesis has not been performed on those residues to study their effects.

Discussion

The region that determines substrate specificity

Plant acyl-ACP TEs can be classified into two families, FatA and FatB (Jones et al. 1995). This study is focused on understanding the substrate specificity of FatB TEs. The FatB class of TEs share high amino acid sequence identity, but have diverse

substrate specificities, suggesting that a small number of residues determine their substrate specificities (Jones et al. 1995). To identify the region(s) that determine the substrate specificity of these enzymes, a systematic domain-shuffling approach was used with two acyl-ACP TEs from the same organism (*C. viscosissima*) that share 70% sequence identity, but express distinct substrate specificities. In two rounds of six-fragment domain shuffling, 18 chimeric TEs were constructed and analyzed. Comparison of the fatty acid products produced by these chimeric TEs demonstrated that substrate specificity is determined by fragments I through IV, which represent the N-terminal two-thirds of the mature proteins. Among these four fragments, fragment III is the most critical determinant, but fragments I, II and IV also affect the substrate specificity, but to smaller degrees. These conclusions appear to differ from those of Yuan et al (Yuan et al. 1995), who concluded that the C-terminal two-thirds of the TE protein is critical for determining substrate specificity. However, when comparing the regions identified by the two studies, there is an approximately 100 amino acid overlap in these two regions (i.e. Fragment III of this study). In fact, Yuan et al. identified that two residues (i.e. M197 and R199 in UcFatB1) affected substrate specificity and the homologous residues reside within the fragment III of CvFatB1, which we show greatly impacts substrate specificity. In another study, characterization of chimeric TEs generated from AtFatA and AtFatB pointed to the N-terminus as a determinant of the substrate specificity (Salas et al. 2002), which is consistent with our results. The predicted tertiary structure of CvFatB2 has two hotdog domains linked by a long coil structure, similar to a previous predicted AtFatB structure (Mayer et al. 2005). The core structure of the N-terminal hotdog consists of fragment II and fragment III. Our results indicate that the N-terminal hotdog domain is the structural basis for the substrate specificity.

Residues determining substrate specificity

Based on the results of the domain-shuffling study and a multiple sequence alignment of 21 functionally characterized plant acyl-ACP TEs, a total of 15 residues in

CvFatB2 sequence were chosen for site-directed mutagenesis analysis to verify their effects on determining the substrate specificity. Comparison of the fatty acid profiles of those site-directed mutants indicated that six residues play critical roles in determining the substrate specificity, including V194 in Fragment II, V217, N223, R226, and R227 in Fragment III, and I268 in Fragment IV. In the multiple sequence alignment, residues R226-R227 follow a conserved Met (M225). In a previous study, mutating the homologous MRR (197-199) in UcFatB1 to RRH changed the substrate specificity of UcFatB1 from 12-carbon fatty acids to producing equal amounts of 12-carbon and 14-carbon fatty acids (Yuan et al. 1995). Besides, mutating the homologous Met in AtFatB to Thr (M174T) also substantially change the specificity toward 16:1 (Mayer et al. 2007). Including this previously tested Met, five residues in the small region from 217 to 227 have been proved to affect specificity significantly. When examining the homologous region (87-97) in the 2OWN structure, we find seven residues in this region are at the bottom of the cavity (Figure 4.9). Our data can be well explained if plant acyl-ACP TE has a similar cavity between the central α-helix and the anti-parallel β-sheets as that in 2OWN. Mutating the critical residues identified herein may change the size of the cavity and thus impact the substrate specificity. For example, mutating V217 and I268 in CvFatB2 to a residue of larger size (Phe) shifted substrate specificity to shorter chain length (Figure 4.7), maybe because the larger residue can make the cavity smaller. Analysis of the tertiary structure 2OWN suggests a few residues in the region from 34 to 45 (homologous to the region from 158 to169 in CvFatB2), which forms the central α-helix, are also important for defining the cavity and thus substrate specificity. Two mutations around this region in AtFatB, M74A and K86Q (corresponding to M158 and K170 in CvFatB2 respectively), were demonstrated to affect the substrate specificity (Mayer et al. 2007), supporting our conclusion from the structure analysis.

Residues that affect the catalytic activity of acyl-ACP TE

In this study, three mutations, L257F, I260L, and L289F, have been proved to increase the total fatty acid production. Western blot analysis of seven representative TE

mutants suggests that the total fatty acid production is mainly affected by the activity of acyl-ACP TE. Therefore, the increased fatty acid production caused by the three mutations indicates these mutations can increase the catalytic activity of acyl-ACP TE. Interestingly, each of these residues (L289, L257, and I260) follow a conserved positively charged residue (either Arg or Lys) in the multiple sequence alignment (Figure S4.1). At between positions 255 and 260, a conserved positively charged motif can be identified: R255(R/K)(L/I/F)S(R/K)(I/M/V/L260). Between positions 285 and 289, there is another positively charged motif, K285(L/V/I)X(K/R)(L/F289). ACP is a small acidic protein. These positively charged motifs may be the binding sites for ACP, and thus are hypothesized as ACP-binding motifs. This hypothesis is similar to a conclusion that ACP binds to a positively charged/hydrophobic region adjacent to the active site tunnel on β-ketoacyl-ACP synthase III (Zhang et al. 2001). Previous structure modeling of *Helianthus annuus* L. acyl-ACP TE HaFatA1 suggested that R212, R213, and K216, which correspond to the positively charged residues at position 255, 256, 259 respectively of CvFatB2, may be involved in the interaction between the enzyme and the substrate acyl-ACP (Serrano-Vega et al. 2005). Substitution at position R212/R213 with either Ala or Glu in HaFatA1 caused dramatic decrease in enzymatic activity, and mutating K216 to Glu significantly increased the Km for 18:1-ACP substrate (Serrano-Vega et al. 2005). Collectively these observations can explain why our mutations L257F, I260L, and L289F increased the activity. The positively charged residues adjacent to these positions may be responsible for the interaction with the acyl-ACP substrate, and the mutations at these positions may change the orientation of surrounding positively charged residues, which in turn affect the interaction with acyl-ACP substrates. In the medium chain specific acyl-ACP TEs (group B), the residues at those positions may have been optimized during evolution to refine the acyl-ACP and TE interactions, which enhances the hydrolysis of medium chain acyl-ACPs, which are less dynamic and more stable than long chain acyl-ACPs (Zornetzer et al. 2010).

A proposed model for substrate specificity

Although there is limited structural and kinetic data concerning acyl-ACP TEs, we can propose a preliminary model based on our modeling and mutagenesis data, integrated with prior studies and analogous characterization of acyl-ACP desaturases that act on the identical substrates, but catalyze acyl-desaturation reactions. Limited kinetic characterizations of both FatA and FatB plant acyl-ACP TEs show that they have similar K_m values, with different substrates of different acyl-chain lengths. But the major difference on their substrate selectivity is the result of different catalytic efficiencies, measured as K_{cat} values (Salas et al. 2002; Serrano-Vega et al. 2005; Sanchez-Garcia et al. 2010; Moreno-Perez et al. 2011). These findings indicate that substrate specificity of different acyl-ACP TEs is more reliant on the reaction velocity, than binding efficiency of the substrates (Serrano-Vega et al. 2005). Analogous conclusions have been drawn from the study of a stearoyl-ACP Δ9 desaturase from *Ricinus communis* (Haas et al. 1999). Analysis of steady-state kinetic parameters of stearoyl-ACP Δ9 desaturase on acyl-ACPs using different acyl chain length substrates indicate that the hydrophobic binding energy plays an important role in the substrate selectivity (Haas et al. 1999). Plant acyl-ACP TEs may have a similar mechanism. The catalytic residues of acyl-ACP TEs that have been identified in Chapter 3 are located close to the mouth of the putative active site cavity. Catalysis is proposed to initiate with the nucleophilic attack of a glutamic acid (D309) carboxylate oxygen anion at the carbonyl carbon of acyl-ACP, which is polarized by hydrogen bonding with the backbone amide group of Asparagine (N316). Prior to the nucleophilic attack, the carboxylate oxygen anion and the carbonyl carbon must take-up a near-attack conformation (Lightstone et al. 1996). The reaction rate is determined by the probability of formation of this near-attack conformation, and the probability of this occurring is dependent upon the free energy of activation (Lightstone et al. 1996). The volume, shape, and hydrophobicity of the cavity can therefore affect the acyl chain binding energy, which may be utilized by the enzyme to position the substrate at the near-attack conformation. For a specific acyl-ACP TE, acyl-ACPs of different chain lengths will have different acyl chain binding energies, which

results in different probability to form near-attack conformation and therefore express different reaction rates.

Acknowledgement

This work was supported by the U.S. National Science Foundation through its Engineering Research Center Program (Award No. EEC-0813570). The authors thank M. Ann D.N. Perera and Zhihong Song of the W.M. Keck Metabolomics Research Laboratory at Iowa State University for assistance with fatty acid analysis.

References

Davies, H. M., Anderson, L., Fan, C. and Hawkins, D. J. (1991). Developmental induction, purification, and further characterization of 12:0-ACP thioesterase from immature cotyledons of Umbellularia californica. Archives of Biochemistry and Biophysics 290(1): 37-45.

Dehesh, K., Jones, A., Knutzon, D. S. and Voelker, T. A. (1996). Production of high levels of 8:0 and 10:0 fatty acids in transgenic canola by overexpression of Ch FatB2, a thioesterase cDNA from Cuphea hookeriana. Plant Journal 9(2): 167-172.

Haas, J. A. and Fox, B. G. (1999). Role of hydrophobic partitioning in substrate selectivity and turnover of the ricinus communis stearoyl acyl carrier protein delta(9) desaturase. Biochemistry 38(39): 12833-12840.

Jing, F., Cantu, D. C., Tvaruzkova, J., Chipman, J. P., Nikolau, B. J., Yandeau-Nelson, M. D. and Reilly, P. J. (2011). Phylogenetic and experimental characterization of an acyl-ACP thioesterase family reveals significant diversity in enzymatic specificity and activity. BMC Biochemistry 12: 44.

Jones, A., Davies, H. M. and Voelker, T. A. (1995). Palmitoyl-Acyl Carrier Protein (Acp) Thioesterase and the Evolutionary Origin of Plant Acyl-Acp Thioesterases. Plant Cell 7(3): 359-371.

Lennen, R. M. and Pfleger, B. F. (2012). Engineering Escherichia coli to synthesize free fatty acids. Trends in Biotechnology 30(12): 659-667.

Lightstone, F. C. and Bruice, T. C. (1996). Ground state conformations and entropic and enthalpic factors in the efficiency of intramolecular and enzymatic reactions .1. Cyclic anhydride formation by substituted glutarates, succinate, and 3,6-endoxo-

Delta(4)-tetrahydrophthalate monophenyl esters. Journal of the American Chemical Society 118(11): 2595-2605.

Liu, X., Sheng, J. and Curtiss, R., 3rd (2011). Fatty acid production in genetically modified cyanobacteria. Proceedings of the National Academy of Sciences of the United States of America 108(17): 6899-6904.

Mayer, K. M. and Shanklin, J. (2005). A structural model of the plant acyl-acyl carrier protein thioesterase FatB comprises two helix/4-stranded sheet domains, the N-terminal domain containing residues that affect specificity and the C-terminal domain containing catalytic residues. Journal of Biological Chemistry 280(5): 3621-3627.

Mayer, K. M. and Shanklin, J. (2007). Identification of amino acid residues involved in substrate specificity of plant acyl-ACP thioesterases using a bioinformatics-guided approach. BMC Plant Biology 7: 1.

Moreno-Perez, A. J., Sanchez-Garcia, A., Salas, J. J., Garces, R. and Martinez-Force, E. (2011). Acyl-ACP thioesterases from macadamia (Macadamia tetraphylla) nuts: cloning, characterization and their impact on oil composition. Plant Physiol Biochem 49(1): 82-87.

Nikolau, B. J., Perera, M. A., Brachova, L. and Shanks, B. (2008). Platform biochemicals for a biorenewable chemical industry. Plant Journal 54(4): 536-545.

Ohlrogge, J. B. and Jaworski, J. G. (1997). Regulation of Fatty Acid Synthesis. Annu Rev Plant Physiol Plant Mol Biol 48: 109-136.

Rawsthorne, S. (2002). Carbon flux and fatty acid synthesis in plants. Progress in Lipid Research 41(2): 182-196.

Salas, J. J. and Ohlrogge, J. B. (2002). Characterization of substrate specificity of plant FatA and FatB acyl-ACP thioesterases. Archives of Biochemistry and Biophysics 403(1): 25-34.

Sanchez-Garcia, A., Moreno-Perez, A. J., Muro-Pastor, A. M., Salas, J. J., Garces, R. and Martinez-Force, E. (2010). Acyl-ACP thioesterases from castor (Ricinus communis L.): an enzymatic system appropriate for high rates of oil synthesis and accumulation. Phytochemistry 71(8-9): 860-869.

Serrano-Vega, M. J., Garces, R. and Martinez-Force, E. (2005). Cloning, characterization and structural model of a FatA-type thioesterase from sunflower seeds (Helianthus annuus L.). Planta 221(6): 868-880.

Steen, E. J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., Del Cardayre, S. B. and Keasling, J. D. (2010). Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. Nature 463(7280): 559-562.

Suzuki, R. and Shimodaira, H. (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics 22(12): 1540-1542.

Thelen, J. J. and Ohlrogge, J. B. (2002). Metabolic engineering of fatty acid biosynthesis in plants. Metabolic Engineering 4(1): 12-21.

Voelker, T. A., Hayes, T. R., Cranmer, A. M., Turner, J. C. and Davies, H. M. (1996). Genetic engineering of a quantitative trait: Metabolic and genetic parameters influencing the accumulation of laurate in rapeseed. Plant Journal 9(2): 229-241.

Voelker, T. A., Worrell, A. C., Anderson, L., Bleibaum, J., Fan, C., Hawkins, D. J., Radke, S. E. and Davies, H. M. (1992). Fatty acid biosynthesis redirected to medium chains in transgenic oilseed plants. Science 257(5066): 72-74.

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association 58(301): 236-&.

White, S. W., Zheng, J., Zhang, Y. M. and Rock (2005). The structural biology of type II fatty acid biosynthesis. Annual Review of Biochemistry 74: 791-831.

Yuan, L., Voelker, T. A. and Hawkins, D. J. (1995). Modification of the substrate specificity of an acyl-acyl carrier protein thioesterase by protein engineering. Proceedings of the National Academy of Sciences of the United States of America 92(23): 10639-10643.

Zhang, X. J., Li, M., Agrawal, A. and San, K. Y. (2011). Efficient free fatty acid production in Escherichia coli using plant acyl-ACP thioesterases. Metabolic Engineering 13(6): 713-722.

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9: 40.

Zhang, Y. M., Rao, M. S., Heath, R. J., Price, A. C., Olson, A. J., Rock, C. O. and White, S. W. (2001). Identification and analysis of the acyl carrier protein (ACP) docking site on beta-ketoacyl-ACP synthase III. Journal of Biological Chemistry 276(11): 8231-8238.

Zornetzer, G. A., Tanem, J., Fox, B. G. and Markley, J. L. (2010). The length of the bound fatty acid influences the dynamics of the acyl carrier protein and the stability of the thioester bond. Biochemistry 49(3): 470-477.

Tables

Table 4.1 Primers for the amplification of six fragments of CvFatB1 and CvFatB2

Fragments	Forward primers	Reverse primers
I	pUC57F(5'-CGGCTCGTATGTTGTGTGGAAT-3')	Cv-R2(5'-GGTACGATCCGCGCCGATTTC-3')
II	Cv-F2(5'-GAAATCGGCGCGGATCGTACC-3')	Cv-R3(5'-CCAGGTCGGATAACGATTGAC-3')
III	Cv-F3(5'-GTCAATCGTTATCCGACCTGG-3')	Cv-R4(5'-CGGGTTTTCTGGTTCATCAT-3')
IV	Cv-F4(5'-ATGATGAACCAGAAAACCCG-3')	Cv-R5(5'-TCGTTCCAACGCGGCGTCAGACC-3')
V	Cv-F5(5'-GGTCTGACGCCGCGTTGGAACGA-3')	Cv-R6(5'-ATCTTCCAGACGCAGCAG-3')
VI	Cv-F6(5'-CTGCTGCGTCTGGAAGAT-3')	pUC57R(5'-CTGCAAGGCGATTAAGTTGGGTAAC-3')

Table 4.2 Pair-wise comparison of TE with three mutations (L257F, I260L, L289F) and TE without triple mutations

TE without three mutations (a)	TE with three mutations (b)	Total FAs (μM)	
		a	b
CvFatB2	CvB2MT14	175.2	108.3
CvB2MT6	CvB2MT17	31.1	113.7
CvB2MT25	CvB2MT34	285.0	464.0
CvB2MT18	CvB2MT33	299.9	569.5
CvB2MT12	CvB2MT29	148.9	287.1
CvB2MT16	CvB2MT32	308.0	330.6
CvB2MT15	CvB2MT31	149.6	393.1
CvB2MT10	CvB2MT27	170.9	230.0
CvB2MT2	CvB2MT22	81.7	84.6
CvB2MT4	CvB2MT23	199.1	384.7
CvB2MT5	CvB2MT24	456.5	400.8
CvB2MT13	CvB2MT28	183.0	328.3
CvB2MT3	CvB2MT21	90.3	201.2

Figures

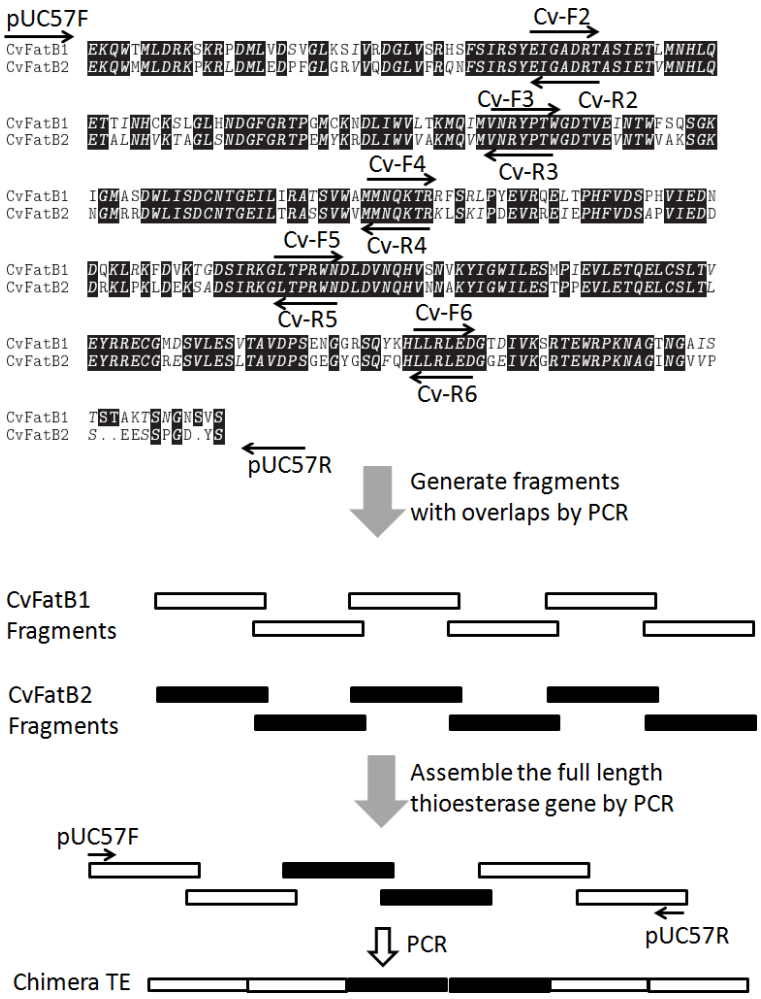


Figure 4.1 Sequence alignment between CvFatB1 and CvFatB2 and schematic diagram for construction of chimeric acyl-ACP TEs. The regions where primers were designed are indicated by arrows.

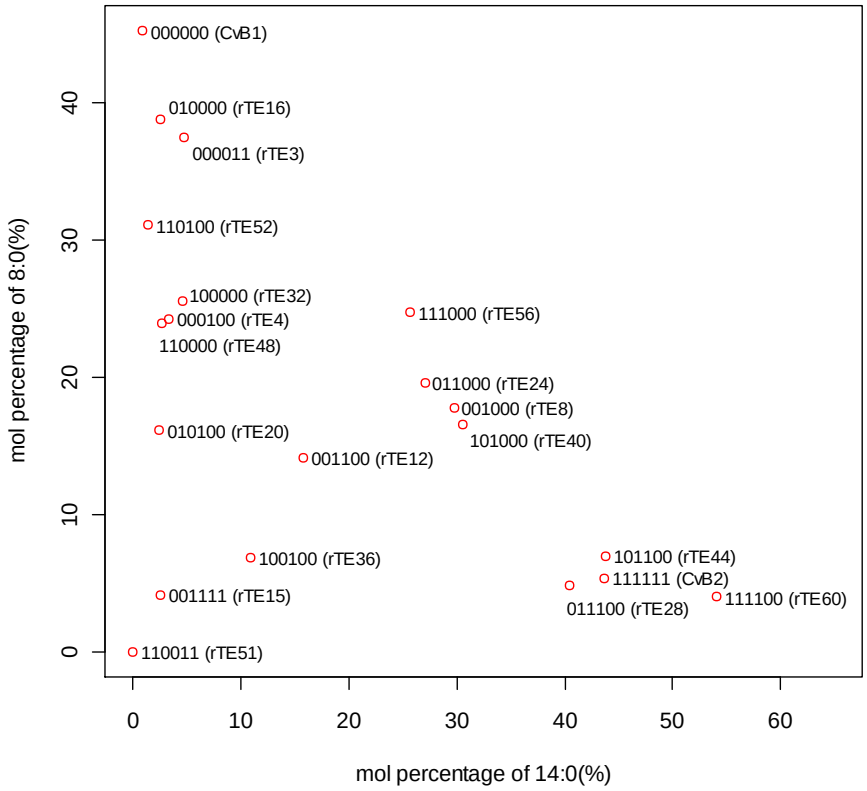


Figure 4.2 Distribution of two major fatty acids produced by chimeric acyl-ACP TEs. Six-digit number indicates the fragments I to VI. "0" indicates a fragment from CvFatB1, while "1" indicates a fragment from CvFatB2. The six-digit binary numbers were converted to decimal numbers which are then used to name the chimeric acyl-ACP TEs.

structure of chimeras						Sample name	total FA (uM)	mol percentage of individual FAs (%)									
I	II	III	IV	V	VI			4:0	6:0	8:0	10:0	10:1	12:0	12:1	14:0	14:1	16:0	16:1
						CvFatB1	541.6	0.0	3.5	45.3	26.1	8.9	3.9	9.6	1.1	0.9	0.0	0.6
						rTE3	511.4	0.0	2.3	42.5	26.6	10.0	4.1	9.5	1.4	1.6	1.3	0.8
						rTE12	49.7	1.2	2.2	18.5	43.0	6.8	2.8	15.6	9.9	0.0	0.0	0.0
						rTE15	56.3	1.3	2.6	4.1	21.0	6.9	14.7	34.3	5.1	10.0	0.0	0.0
						rTE48	812.3	2.6	12.2	23.8	17.8	7.2	8.9	18.3	1.7	5.8	0.8	0.8
						rTE51	6.4	4.4	1.9	0.0	20.9	16.4	2.4	30.0	21.8	2.1	0.0	0.0
						rTE60	142.2	0.2	0.5	3.6	1.0	0.1	1.1	0.9	54.9	0.1	10.1	27.4
						CvFatB2	175.2	0.0	0.3	5.4	5.4	1.4	1.3	3.2	45.2	0.6	5.5	31.7

Figure 4.3 Substrate specificities of wild-type and chimeric acyl-ACP TEs in the first round of domain-shuffling. The sequence structures of these TEs are presented as different color combinations. The sequences fromCvFatB1 are shown in orange and the sequences from CvFatB2 in blue. The major fatty acids produced by acyl-ACP TEs are highlighted in green. The intensity of green color indicates the relative abundance.

structure of chimeras						Sample name	total FA (uM)	mol percentage of individual FAs (%)										
I	II	III	IV	V	VI			4:0	6:0	8:0	10:0	10:1	12:0	12:1	14:0	14:1	16:0	16:1
						CvFatB1	541.6	0.0	3.5	45.3	26.1	8.9	3.9	9.6	1.1	0.9	0.0	0.6
						rTE4	491.2	0.1	1.1	24.3	36.0	11.9	4.1	12.2	3.6	3.8	0.0	3.0
						rTE8	87.1	0.4	4.9	17.9	10.2	2.0	3.4	5.3	31.6	6.4	0.0	17.9
						rTE16	820.3	0.4	1.4	38.9	15.3	6.6	9.3	16.2	2.7	7.5	0.0	1.8
						rTE20	387.0	0.2	0.7	16.1	19.9	8.1	12.4	24.3	2.8	13.2	0.0	2.2
						rTE24	276.2	1.2	1.2	19.8	4.1	1.3	6.2	7.3	27.8	6.2	1.0	23.9
						rTE28	279.8	0.3	0.4	4.8	1.0	0.2	1.3	1.7	41.4	1.8	2.3	44.8
						rTE32	488.7	0.1	1.3	25.5	24.6	8.2	11.0	16.1	4.8	4.2	0.0	4.3
						rTE36	192.1	0.5	0.9	6.9	18.5	4.3	13.1	17.6	11.7	17.4	0.0	9.2
						rTE40	136.1	0.7	5.7	16.7	3.1	0.3	6.0	4.7	32.0	10.6	0.0	20.1
						rTE44	46.9	1.5	4.2	6.8	1.7	0.0	1.3	1.1	46.4	5.8	0.0	31.3
						rTE52	874.3	7.2	17.4	31.0	11.4	8.1	5.3	12.3	1.6	4.6	0.0	1.0
						rTE56	405.2	2.8	2.2	24.9	5.6	2.4	7.5	8.2	26.2	4.6	0.8	14.8
						rTE60	142.2	0.2	0.5	3.6	1.0	0.1	1.1	0.9	54.9	0.1	10.1	27.4

Figure 4.4 Substrate specificities of wild-type and chimeric acyl-ACP TEs in the second round of domain-shuffling. The sequence structures of these TEs are presented as different color combinations. The sequences fromCvFatB1 are shown in orange and the sequences from CvFatB2 in blue. The major fatty acids produced by acyl-ACP TEs are highlighted in green. The intensity of green color indicates the relative abundance.

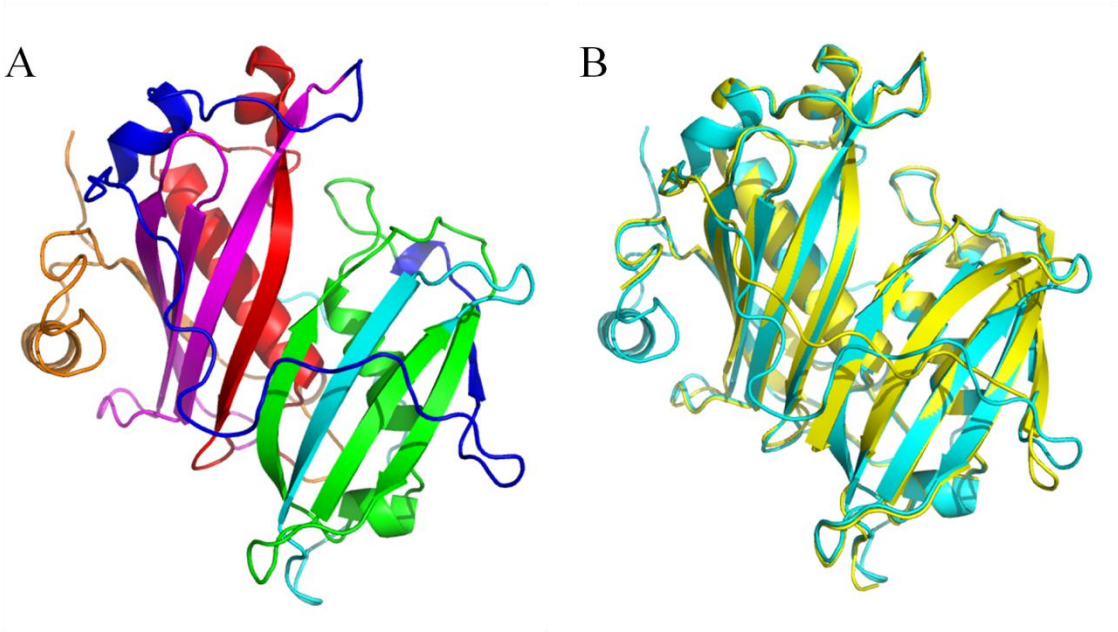


Figure 4.5 Predicted structure of CvFatB2 (A) and structure alignment of CvFatB2 and 2OWN (B). A, Fragments I to VI are shown in orange, red, magenta, blue, green, and cyan respectively; B, structure of CvFatB2 is shown in cyan and 2OWN in yellow.

		194	217	223	257	260	268	276	282	289
Group A	AEM72523 CvFatB2	V	V	N	L	I	I	A	D	L
	AEM72524 CvFatB3	V	V	N	L	I	I	A	D	L
	AAC49180 CpFatB2	V	V	H	L	I	I	A	D	L
	AAB51525 GmFatB1	V	V	N	L	I	I	D	D	L
	EER87824 SbFatB1	V	V	N	L	M	I	S	Q	P
	EER88593 SbFatB2	V	V	N	L	I	I	S	D	L
	AEM72519 CnFatB1	V	V	N	L	M	I	A	D	L
	AAD42220 EgFatB	V	V	N	L	M	I	A	D	L
	AEM72520 CnFatB2	V	V	N	L	V	I	A	D	L
	AAG43857 IgFatB1	V	A	N	L	F	I	V	D	L
	AAG43858 IgFatB2	V	A	N	L	F	I	A	D	L
	EDQ65090 PpFatB	V	V	N	L	M	I	F	M	L
	EER96252 SbFatB3	V	V	N	L	I	V	S	S	L
	AAC48990 ChFatB1	V	V	N	L	I	I	P	D	L
Group B	AEM72521 CnFatB3	V	I	N	L	L	I	D	N	F
	AAB71731 UaFatB1	V	A	N	I	I	I	P	D	F
	AAC49269 ChFatB2	V	F	I	L	L	I	P	S	F
	CAB60830 ClFatB3	L	F	I	F	L	L	P	N	F
	AAC49784 CwFatB2	L	F	I	F	L	I	P	N	F
	AAC49179 CpFatB1	V	L	I	F	L	F	P	N	F
	AEM72522 CvFatB1	L	F	I	F	L	L	P	N	F

Figure 4.6 Part of the multiple sequence alignment. Numbers of the residues are according to the sequence of CvFatB2. Cv, *Cuphea viscosissima*; Cp, *Cuphea palustris*; Gm, *Garcinia mangostana*; Sb, *Sorghum bicolor*; Cn, *Cocos nucifera*; Eg, *Elaeis guineensis*; Ig, *Iris germanica*; Pp, *Physcomitrella patens*; Ch, *Cuphea hookeriana*; Ua, *Ulmus americana*; Cl, *Cuphea lanceolata*; Cw, *Cuphea wrightii*.

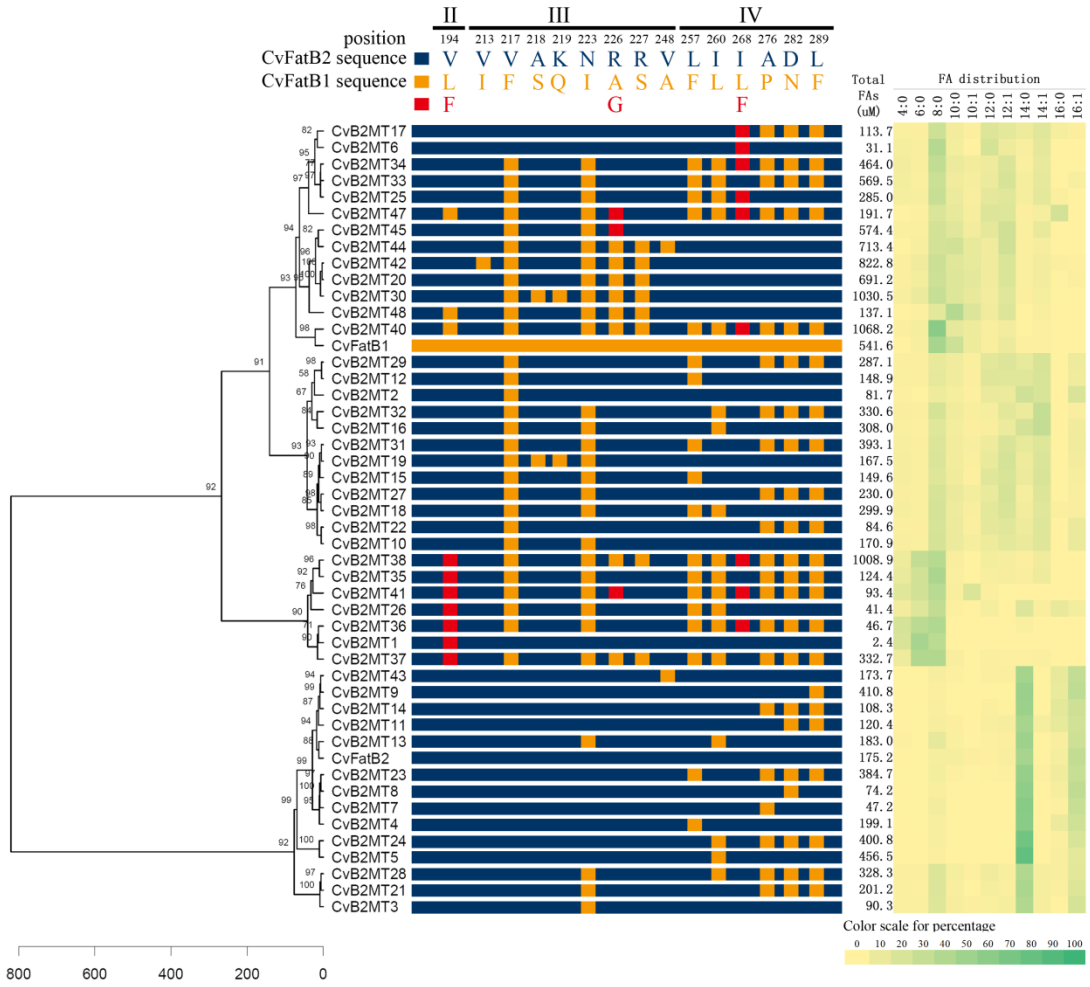


Figure 4.7 Fatty acid profiles and cluster analysis of CvFatB2 point mutants. Wild-type CvFatB1 and CvFatB2 are shown as blue and orange lines respectively. Schematic graph in the middle shows the point mutations for each mutant. Blue color represents sequences from CvFatB2, orange color represents sequences from CvFatB1, and red color represents residues not from CvFatB1 or CvFatB2. The color coded fatty acid profile is shown on the right side. Major fatty acids produced by each mutant were highlighted in green. Left side is a cluster analysis based on the fatty acids profile.

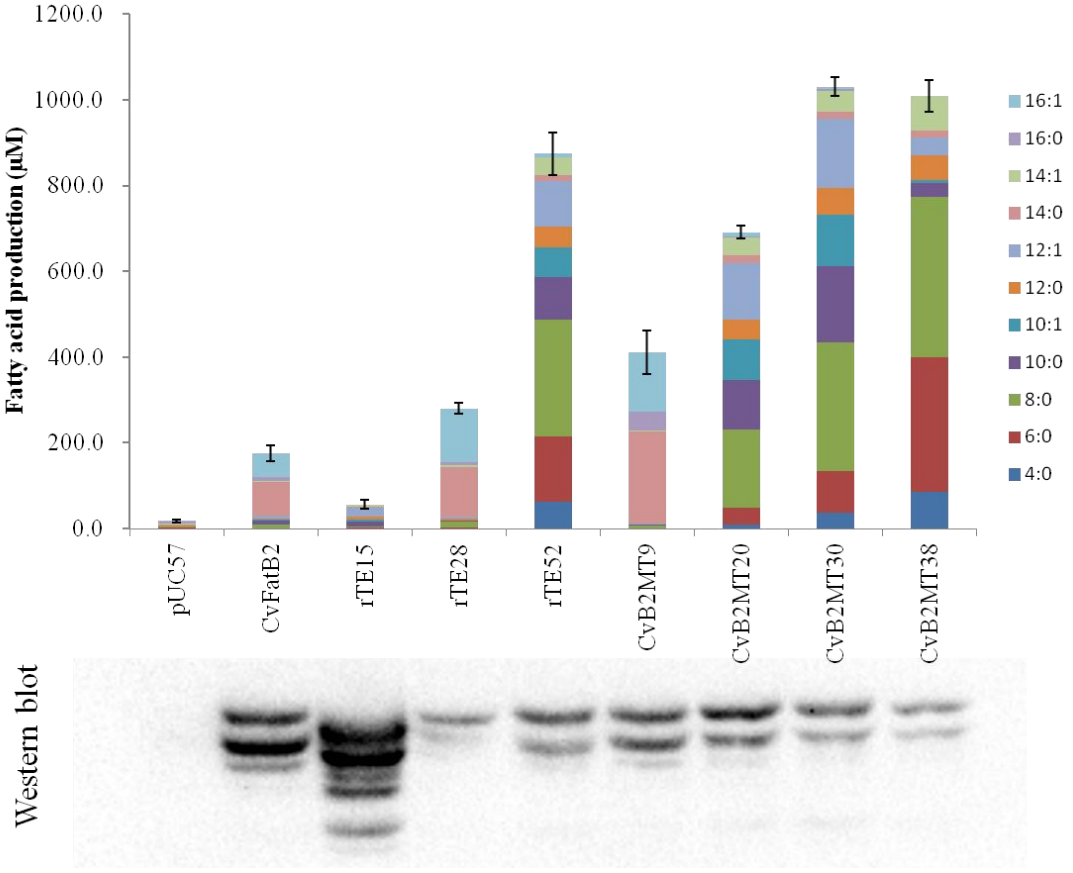


Figure 4.8 Fatty acid production and western blot analysis of wild-type CvFatB2 and seven representative mutants. Fatty acid production is an average of 4 replicates. Error bars represent standard error of mean.

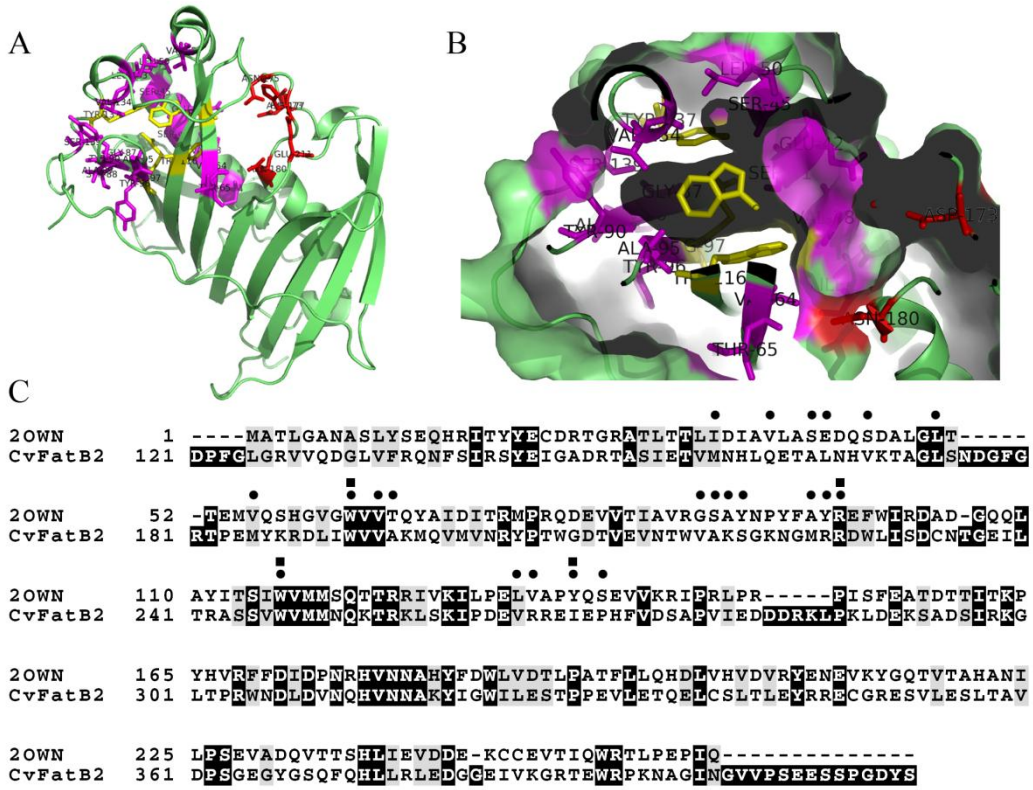


Figure 4.9 Structural analysis of 2OWN (A and B) and sequence alignment of 2OWN and CvFatB2. A, cartoon view of crystal structure 2OWN. Catalytic resiudes are shown in red stick models. The residues that form an active site cavity in the N-terminal hotdog domain are shown in magenta and yellow stick models. The four bulky residues that occupy more than 50% of the cavity surface area are shown in yellow stick models. B, active site cavity in the N-terminal hotdog domain. Same color scheme was used as in A. C, sequence alignment of 2OWN and CvFatB2. Residues forming the cavity are indicated by dots and the bulky residues are indicated by squares.

Supplemental material

Figure S4.1

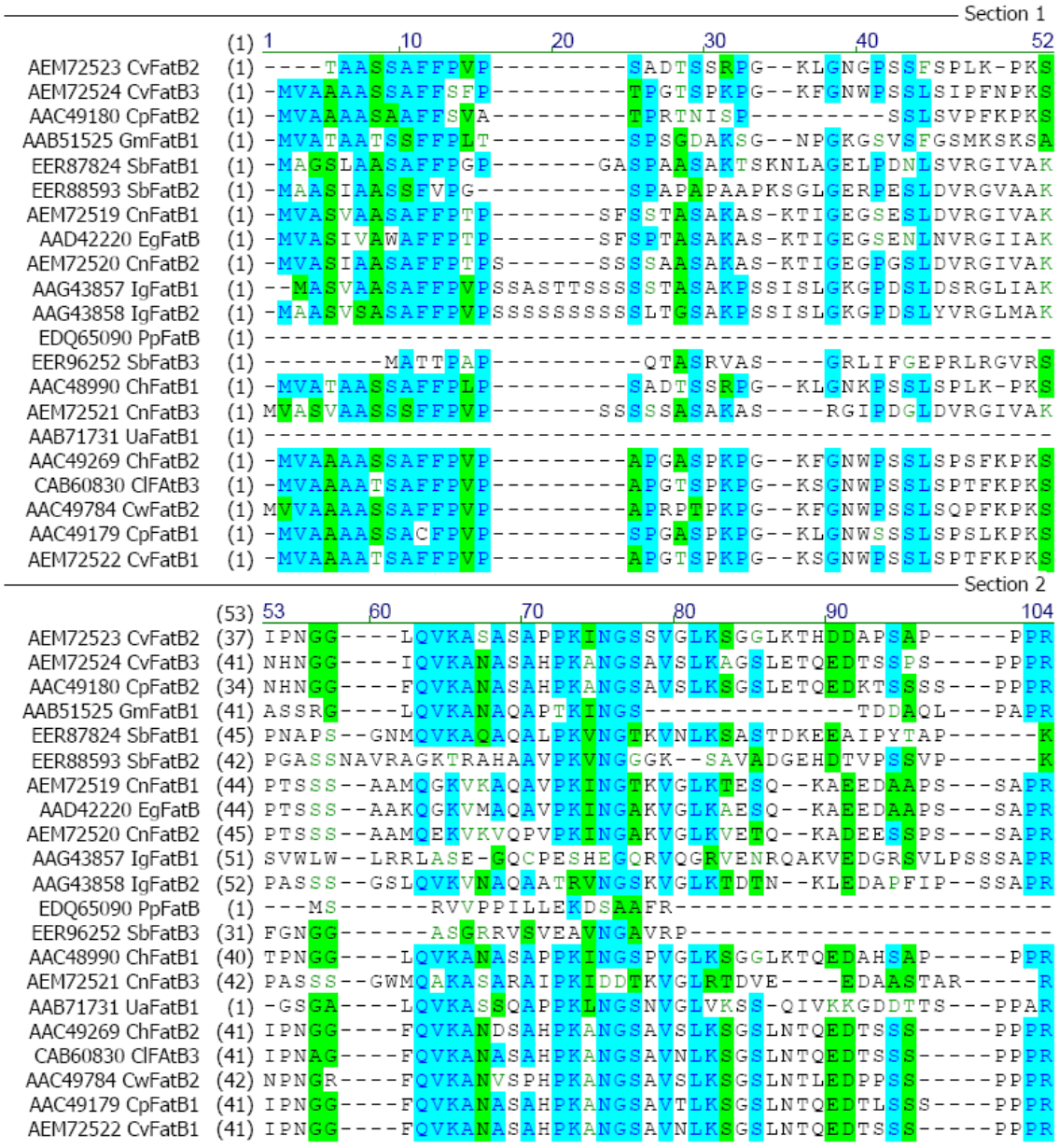


Figure S4.1 continued

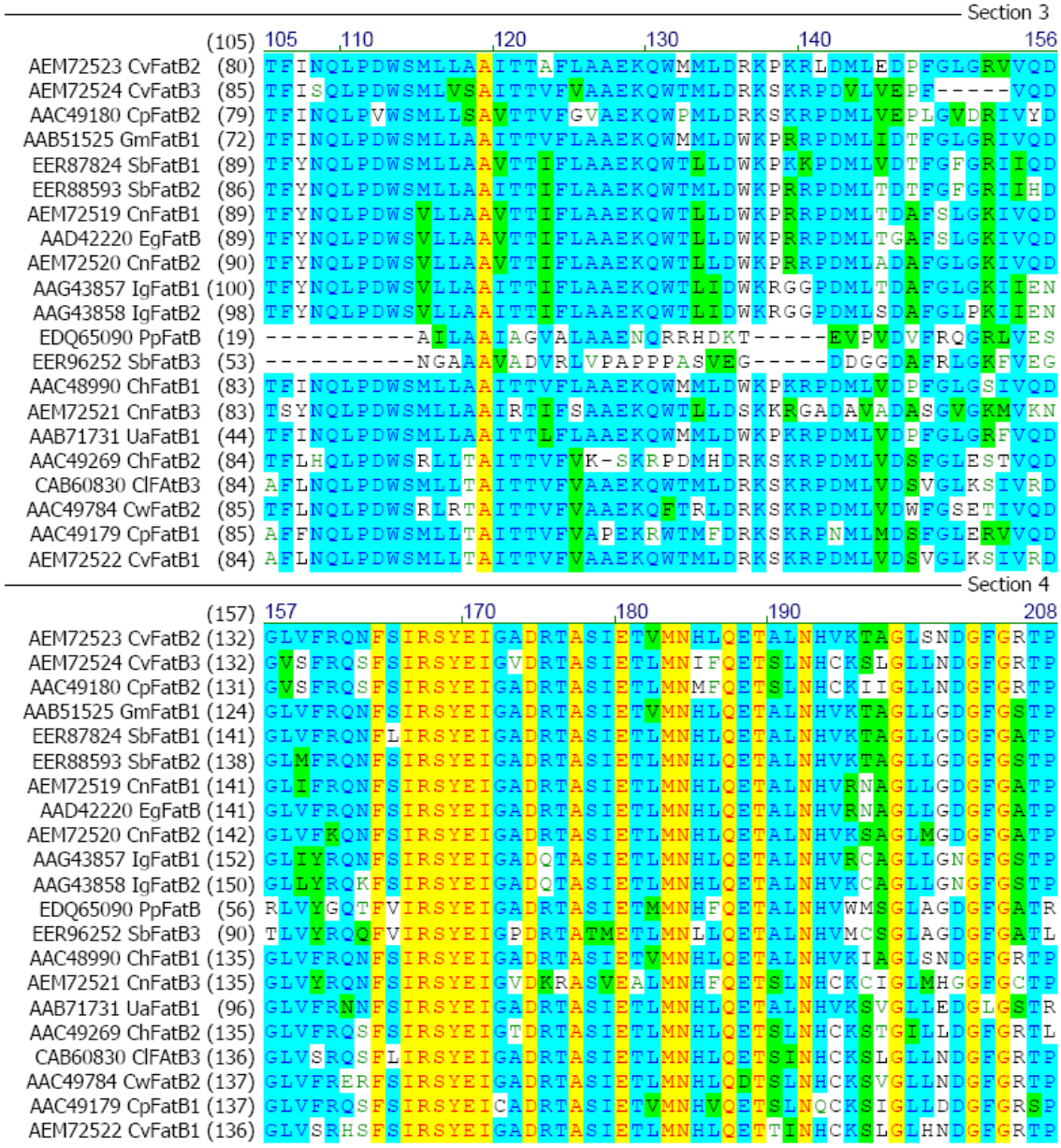


Figure S4.1 continued

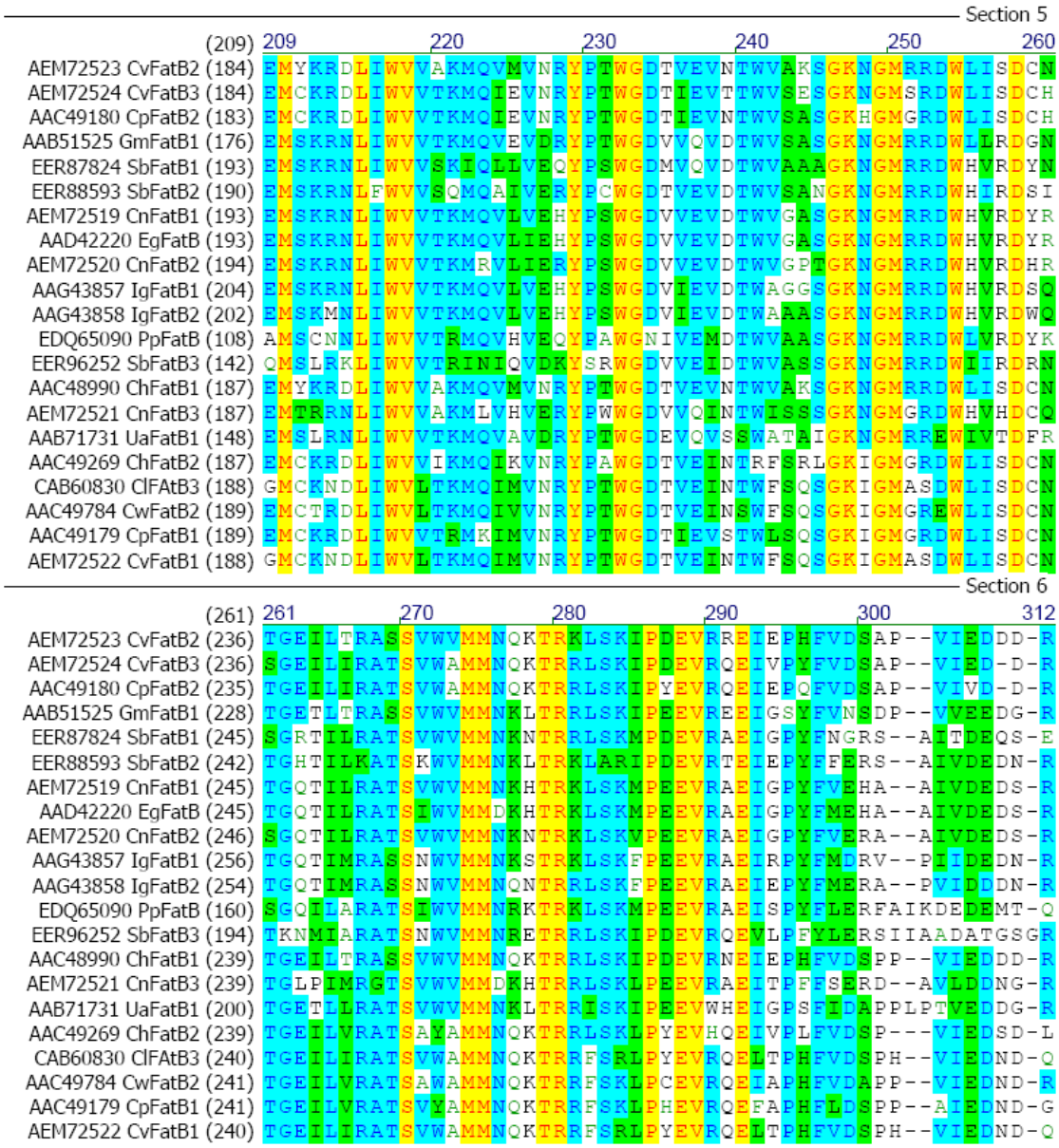


Figure S4.1 continued

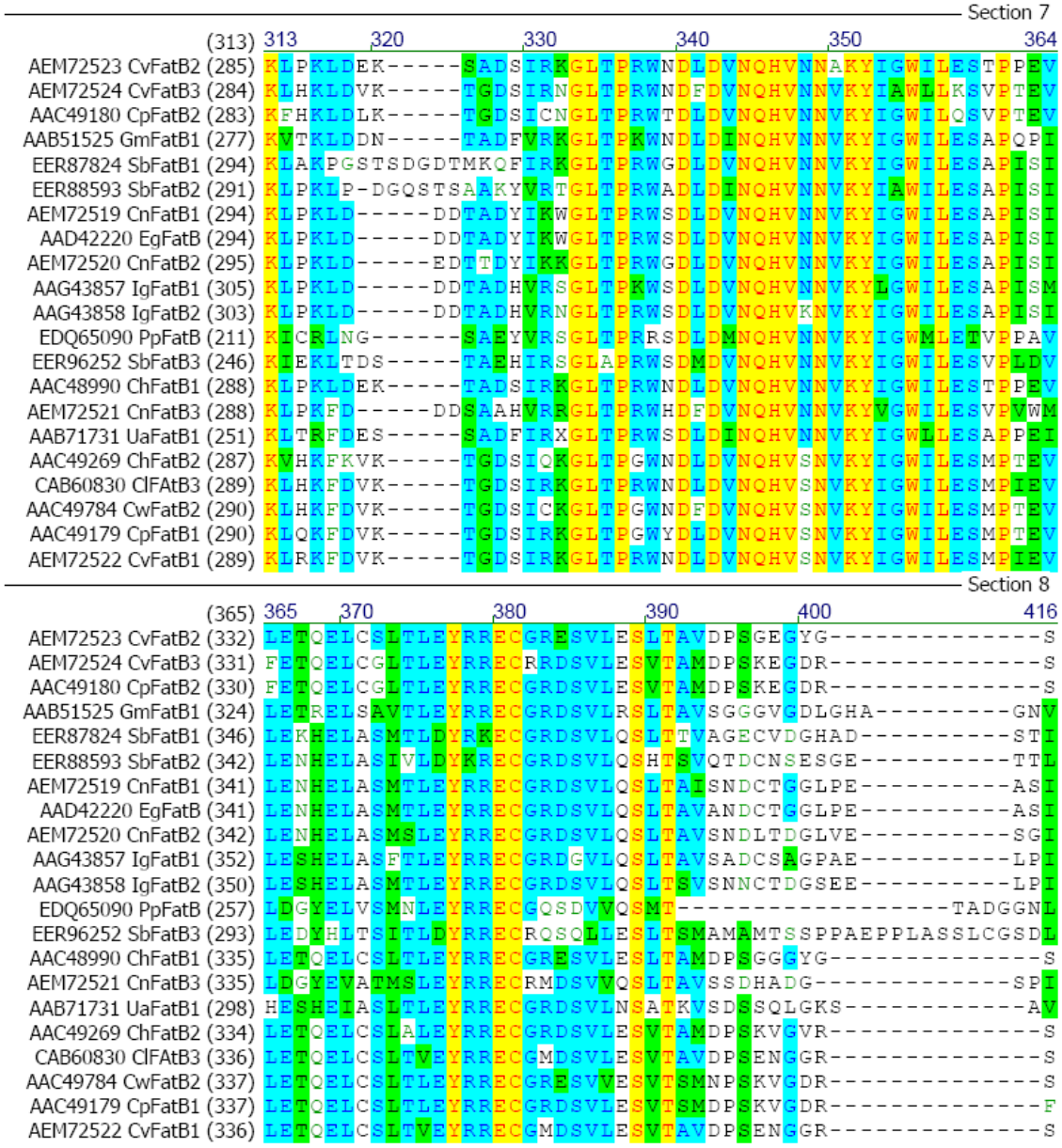


Figure S4.1 continued

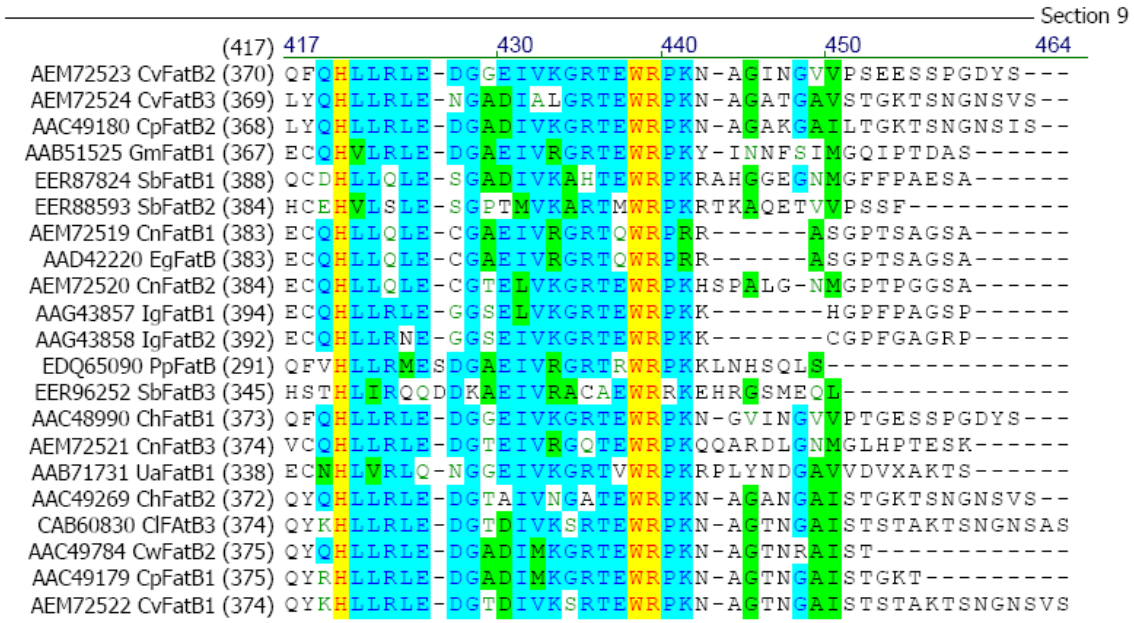


Figure S4.1 Multiple sequence alignment of 21 functionally characterized plant acyl-ACP TEs.

Sample name	mol percentage (%)											total FAs (μM)
	4:0	6:0	8:0	10:0	10:1	12:0	12:1	14:0	14:1	16:0	16:1	
CvB2MT17	2.0	1.7	26.2	6.5	1.5	17.6	13.5	8.6	15.8	5.0	1.6	113.7
CvB2MT6	4.6	3.1	38.2	5.8	0.2	21.7	9.1	3.4	13.9	0.0	0.0	31.1
CvB2MT34	6.1	3.9	31.8	9.6	6.3	9.2	18.4	2.8	11.2	0.3	0.4	464.0
CvB2MT33	5.8	2.5	27.2	7.9	5.3	9.6	20.4	5.1	12.5	0.0	3.6	569.5
CvB2MT25	4.5	3.6	28.8	9.3	3.4	13.0	15.9	5.2	13.7	2.0	0.6	285.0
CvB2MT47	0.6	0.9	18.5	12.0	4.2	17.9	17.5	1.4	4.5	22.6	0.0	191.7
CvB2MT45	1.9	2.0	22.5	15.8	6.2	13.1	24.4	2.6	11.4	0.0	0.0	574.4
CvB2MT44	0.7	3.5	22.7	23.6	11.3	7.9	24.2	1.8	4.0	0.0	0.2	713.4
CvB2MT42	1.8	7.1	27.0	15.0	12.1	6.9	18.9	2.6	6.1	1.5	1.1	822.8
CvB2MT20	1.5	5.6	26.2	16.9	13.6	6.6	19.1	2.4	6.2	0.3	1.4	691.2
CvB2MT30	3.5	9.5	29.2	17.2	11.7	6.1	15.7	1.5	4.7	0.4	0.6	1030.5
CvB2MT48	0.1	1.2	13.7	38.6	18.1	6.7	17.5	0.0	3.9	0.2	0.0	137.1
CvB2MT40	1.8	4.8	61.4	12.5	8.5	3.3	6.2	0.6	0.9	0.0	0.0	1068.2
CvFatB1	0.0	3.5	45.3	26.1	8.9	3.9	9.6	1.1	0.9	0.0	0.6	541.6
CvB2MT29	2.4	1.5	11.4	7.2	2.5	15.0	18.1	12.1	20.5	0.4	8.9	287.1
CvB2MT12	2.2	1.6	12.5	7.2	1.6	16.4	15.4	15.3	20.0	0.0	7.8	148.9
CvB2MT2	1.0	0.7	8.9	4.6	1.1	10.0	9.7	24.5	19.9	0.1	19.5	81.7
CvB2MT32	3.1	1.3	21.5	7.3	3.9	4.8	14.4	9.5	30.6	0.0	3.4	330.6
CvB2MT16	2.2	0.9	15.4	5.6	1.6	3.9	9.6	22.8	30.9	0.0	7.0	308.0
CvB2MT31	3.4	2.2	17.9	5.9	3.7	9.0	19.6	12.1	17.3	0.7	8.1	393.1
CvB2MT19	4.4	2.1	18.2	6.3	2.8	10.3	22.0	8.5	18.4	0.6	6.3	167.5
CvB2MT15	4.0	2.1	16.1	7.5	1.2	15.6	19.8	9.2	19.8	0.7	4.0	149.6
CvB2MT27	3.3	2.0	20.8	9.9	5.8	10.3	21.5	5.3	17.6	0.3	3.2	230.0
CvB2MT18	4.2	1.9	21.9	8.4	3.1	13.6	19.9	6.1	15.3	2.3	3.2	299.9
CvB2MT22	2.9	1.6	20.7	9.2	1.5	13.0	15.6	8.9	16.7	0.0	10.0	84.6
CvB2MT10	3.0	2.2	21.6	9.6	4.6	8.8	14.1	9.8	16.4	0.0	9.9	170.9
CvB2MT38	8.5	31.1	37.2	3.1	0.6	5.7	4.2	1.6	8.1	0.0	0.0	1008.9
CvB2MT35	10.5	23.6	43.1	2.5	0.6	3.6	4.1	3.0	8.8	0.0	0.1	124.4
CvB2MT41	15.1	23.9	36.2	3.8	18.9	1.4	0.7	0.0	0.0	0.0	0.0	93.4
CvB2MT26	6.3	17.2	31.7	1.2	0.0	4.8	0.0	17.9	3.9	11.3	5.7	41.4
CvB2MT36	22.1	35.2	40.3	0.0	0.0	0.0	0.0	2.4	0.0	0.0	0.0	46.7
CvB2MT1	19.4	46.0	34.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.4
CvB2MT37	8.9	40.1	40.2	3.4	0.3	2.8	1.0	1.1	2.0	0.0	0.2	332.7
CvB2MT43	0.4	0.7	4.1	1.2	0.5	1.5	1.4	47.0	0.7	11.5	31.0	173.7
CvB2MT9	0.1	0.1	1.5	0.3	0.1	0.3	0.4	52.6	0.6	10.4	33.7	410.8
CvB2MT14	0.0	0.4	3.4	0.8	0.0	1.6	0.8	52.2	0.7	14.1	26.0	108.3
CvB2MT11	0.1	0.4	4.3	1.5	0.2	3.8	4.0	45.5	4.9	11.9	23.5	120.4
CvB2MT13	1.0	0.8	12.7	2.1	0.7	0.9	1.5	50.7	0.5	3.1	26.0	183.0
CvFatB2	0.0	0.3	5.4	5.4	1.4	1.3	3.2	45.2	0.6	5.5	31.7	175.2
CvB2MT23	0.6	0.4	5.6	1.0	0.3	0.8	0.8	54.2	0.9	6.6	28.7	384.7
CvB2MT8	0.2	0.4	3.9	0.3	0.0	0.0	0.0	58.0	0.0	4.4	32.8	74.2
CvB2MT7	0.7	0.7	6.1	0.6	0.0	0.0	0.0	62.2	0.0	0.9	28.8	47.2
CvB2MT4	0.2	0.3	3.7	0.6	0.1	0.0	0.2	60.9	0.0	9.3	24.5	199.1
CvB2MT24	0.2	0.1	3.6	0.4	0.1	0.0	0.3	73.6	0.2	1.4	20.0	400.8
CvB2MT5	0.0	0.1	1.5	0.2	0.0	0.0	0.0	81.5	0.0	2.9	13.9	456.5
CvB2MT28	1.7	1.4	17.6	5.0	1.2	7.5	3.7	40.1	4.1	3.4	14.3	328.3
CvB2MT21	1.9	1.1	17.8	3.4	0.6	4.6	2.7	43.8	4.5	2.0	17.5	201.2
CvB2MT3	2.6	1.9	17.2	4.1	1.5	2.1	4.6	40.1	4.3	0.0	21.6	90.3

Figure S4.2 Fatty acid production and profiles of site-directed mutants. Major fatty acids produced by the mutants are highlighted in green.

CHAPTER 5 – DIRECTED EVOLUTION OF ACYL-ACP THIOESTERASE FOR HIGHER FATTY ACID PRODUCTION

Fuyuan Jing, Marna D. Yandeau-Nelson, Basil J. Nikolau
Department of Biochemistry, Biophysics, and Molecular Biology, Iowa State University

A manuscript to be submitted for publication

Abstract

Acyl-ACP TE, controlling the terminal hydrolysis reaction of acyl-ACPs, plays an essential role in determining the fatty acid production and composition of the fatty acid synthesis (FAS) pathway. Acyl-ACP TE is of great biotechnological interest because it has recently been widely used in the metabolic engineering of the FAS pathway in microbes or algae for the production of fatty acids as biofuels or biorenewable chemicals. A directed evolution approach was developed in this study to improve the fatty acid productivity of acyl-ACP TE. The variant library was designed based on the multiple sequence alignment of six plant acyl-ACP TEs, which exhibit different fatty acid productivity and substrate specificities. This library was generated by assembling 30 degenerate oligos, expressed in *E. coli* strain K27, and screened on Neutral Red plates. Analysis of the fatty acid productivity of variants displaying different intensities of red colony colors with GC-MS demonstrated that more intensive red color of the colony is an indicator for high TE activity of the variants. Screening the library on Neutral Red plates and analyzing the dark red colonies with GC-MS resulted in identification of TE variants with increased fatty acid productivity. Among them, one variant, TEGm162, was recovered with very high frequency, which produced 30% more total fatty acids than most productive parental TE and 10-fold more than the least productive parental TE. Comparison of the sequences and substrate specificities of the TE variants led to the identification of a region that may be involved in determining the substrate specificity of acyl-ACP TE, demonstrating that directed evolution is a powerful tool to study the sequence-specificity relationship of acyl-ACP TE.

Introduction

Fatty acids, naturally stored in the form of triacylglycerol in plant seeds oil and animal fats, are very abundant renewable resources. They are consumed by humans not only in the food supply but also as industrial feedstocks. Fatty acids are ingredients of many products, such as soaps, detergents, surfactants, lubricants, cosmetics, and pharmaceuticals (Ohlrogge 1994; Thelen et al. 2002; Dyer et al. 2008). In recent years, the use of fatty acids for the production of biofuels or chemical feedstocks has been increasing, mostly driven by the rapidly rising cost of petroleum and growing environmental concerns about consuming large amounts of fossil fuel (Durrett et al. 2008; Nikolau et al. 2008).

In both plants and bacteria, fatty acids are synthesized via type II fatty acid synthase. This pathway starts with the condensation of malonyl-ACP and acetyl-CoA catalyzed by a ketoacyl-ACP synthase III (KAS III) to generate acetoacetyl-ACP, followed by three consecutive reduction reactions to yield a butyl-ACP. The following condensation reactions of malonyl-ACP and acyl-ACPs are catalyzed by KAS I or KAS II enzymes. The condensation reduction cycle can repeat several times with addition of two-carbon atoms to the acyl chain length for each cycle. The elongation process usually ends at 16-18 carbon acyl-ACPs, and the resulting fatty acids can be transacylated to a glycerol backbone to synthesize membrane lipids in bacteria, or in plant plastids acyl-ACP thioesterase intercepts the elongation cycle by hydrolyzing the thioester bond of acyl-ACPs to release the free fatty acids and ACP. Many plant acyl-ACP TEs have been isolated, characterized, and demonstrated to play essential roles in determining the fatty acid chain length of storage lipids (Pollard et al. 1991; Voelker et al. 1992; Leonard et al. 1998).

Fatty acids are normally obtained from plant oil and animal fats. Specifically, four oil crops, namely oil palm, soybeans, rapeseed, and sunflowers, supply about 79% of the world's total oil production (Dyer et al. 2008). With the better understanding of the regulation of fatty acid biosynthesis pathway in bacteria, interest has developed in the possibility of producing fatty acids in an alternative way, which may pave the way

for wide commercialization of fatty acid products (Magnuson et al. 1993; Handke et al. 2011). Intensive interests have been shown in the metabolic engineering of fatty acid biosynthesis pathway in microbes or algae for the production of free fatty acids or fatty acid derivatives (Liu et al. 2011; Zhang et al. 2011; Lennen et al. 2012; Ranganathan et al. 2012; Zhang et al. 2012). These efforts have primarily focused on improving free fatty acid production in microbes, mainly *Escherichia coli*, and two approaches have been demonstrated to be the most critical steps (Lennen et al. 2012). One is the overexpression of acyl-ACP TE to release fatty acids from the FAS pathway, and another is impairing the fatty acid β-oxidation pathway with fadD or fadE mutations to prevent the reuse of the free fatty acids. Acyl-ACP TEs not only control the fatty acid composition by selectively hydrolyzing the acyl-ACPs (Voelker et al. 1994), but also enhance fatty acid production through two effects: (1) they create a new product sink; and (2) they elevate the metabolic flux through the FAS pathway by depleting long chain acyl-ACPs, which act as feedback inhibitors for upstream enzymes in the FAS pathway (Jiang et al. 1994; Heath et al. 1996; Lennen et al. 2012). Optimization of the expression level of acyl-ACP TEs with different plasmid copy numbers or different promoter strengths, leads to significant improvements of free fatty acid production in *E. coli*, demonstrating the crucial role of acyl-ACP TE in metabolic engineering of FAS pathway for the production of these molecules (Steen et al. 2010; Youngquist et al. 2012; Zhang et al. 2012). All the acyl-ACP TEs used in these previous studies are from a variety of different natural resources, each of which have evolved for that organism's environmental niche. But that natural evolutionary adaptation may not be optimal for the envisioned industrial application in an *E. coli* strain bioengineered for fatty acid productivity. This study was undertaken to optimize the catalytic efficiency of acyl-ACP TE with the goal of improving fatty acid productivity in microbes.

Limited structural information and a not well-understood catalytic mechanism for acyl-ACP TEs make it difficult to increase the activity of this enzyme by rational design. Directed evolution, which mimics the natural evolutionary process at the lab scales, is a promising alternative approach (Dougherty et al. 2009; Turner 2009; Cobb et al. 2013).

Directed evolution involves iterative rounds of random mutagenesis and screening for the desired properties. It has been successfully applied on many biocatalysts to tailor their functions, including substrate specificity, catalytic turn-over, and thermostability (Nair et al. 2008; Zha et al. 2008; Turner 2009). Because it is possible to screen for higher total fatty acid production, considering fatty acids have a carboxylic acid group with pK_a ~4.9 we adapted a screening method for TE activity (Mayer et al. 2007), and developed a platform for the directed evolution of acyl-ACP TEs for higher fatty acid productivity in *E. coli.* Starting with six functionally characterized plant acyl-ACP TEs that have diverse protein sequences, catalytic efficiencies and substrate specificities (Jing et al. 2011), we designed and generated a mutant library with a theoretical 10^{38} possible TE variants. Screening this library for TEs that express higher fatty acid productivity enabled the isolation and characterization of TE variants that show higher catalytic efficiency.

Methods

Design of the primers

Protein sequences of the six plant acyl-ACP TEs that were used in this study are: CvFatB1 and CvFatB2 from *Cuphea viscosissima*, CnFatB2 and CnFatB3 from *Cocos nucifera*, UaFatB1 from *Ulmus americana*, and CpFatB1 from *Cuphea hookeriana.* These sequences were subjected to multiple sequence alignment with ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) (Figure 5.1). The N-terminal chloroplastic transit peptide and the proposed hydrophobic membrane anchor (Facciotti et al. 1998), which can be experimentally removed without affecting TE activity and specificity of acyl-ACP TEs, were not included in this characterization. The conserved and variable residues were identified from the multiple sequence alignment. Among the 307 residue positions, 65 positions were chosen for random mutagenesis with 2-8 possible substitutions at each position, and consensus sequences were used for other positions. The protein sequence was back-translated into nucleotide sequence with the GeneDesign

algorithm (http://54.235.254.95/gd/) (Richardson et al. 2006). A total of 30 DNA oligos were designed to assemble the TE mutant library, with mixed nucleotides incorporated into the corresponding positions for variable residues (Table 5.1). Each of these oligos have 22-25 nucleotide overlaps with each other, with Tm values of 54-56℃. Restriction sites of BamHI and EcoRI were incorporated into oligos M1-1 and M3-10 respectively. Theoretically, the maximum number of variants in this mutant library is up to 10^{38}.

Generation of the mutant library

The acyl-ACP TE mutant library was generated by assembling 30 oligo primers in two rounds of PCR. The first round of PCR was conducted in 50 μL of reaction mixture containing 0.15 μM of each primer, 1 x Taq PCR buffer, 0.4 mM dNTP, 3 mM $MgCl_2$, and 1 Unit of Taq DNA polymerase (New England Biolabs, USA) using a cycling program of 95 ℃ for 3 min, 55 cycles of 95 ℃ for 15 s, 50 ℃ for 20 s and 68 ℃ for 40 s, and a final extension step of 68 ℃ for 5 min. The second round of PCR was performed in 8 tubes of 50 μL of reaction mixture which contained 1 x Taq PCR buffer, 2 μL of first round PCR product as template, 1.5 mM $MgCl_2$, 0.2 mM dNTP, 0.2 μM M1-1 and M3-10 primers each, and 1 Unit Taq DNA polymerase using a cycling program of 95 ℃ for 3 min, 28 cycles of 95 ℃ for 15 s,60 ℃ for 20 s and 68 ℃ for 40 s, and a final extension step of 68 ℃ for 5 min. The second round PCR products were pooled, separated by electrophoresis on 1% agarose gel, purified with the QiaQuick gel extraction kit (Qiagen, Valencia, CA, USA), digested with BamHI and EcoRI, and cloned into the corresponding restriction sites of the pUCHisGm vector (Figure 5.2). The pUCHisGm vector is modified from pUC57 by the insertion of a 6x His tag at the N-terminus of the cloned TE, and fusion of a gentamicin resistant gene (Gm^R) at the C-terminus of the cloned TE gene. When cloned into pUCHisGm vector, the acyl-ACP TE variant gene was fused to Gm^R gene with a 3 x GGGS flexible linker (Chen et al. 2012) between them and transcriptionally controlled by a weak lacZ promoter. The constructed vectors containing the variant genes were transformed into *E. coli* K27 by electroporation. The six wild-type acyl-ACP TEs (UaFatB1, CpFatB1, CvFatB1,

CvFatB2, CnFatB2, and CnFatB3) were also cloned into pUCHisGm and transformed into *E coli* strain K27 as controls.

Screening for TE activity on the Neutral Red Plates

The initial screening of the TE variants was conducted on Neutral Red plates, which were M9 minimal medium (47.7 mM Na_2HPO_4, 22.1 mM KH_2PO_4, 8.6 mM NaCl, 18.7 mM NH_4Cl, 2 mM $MgSO_4$, and 0.1 mM $CaCl_2$) solidified by 15 g/L agar and supplemented with 0.4% glucose, 100 mg/L carbenicillin, 2.5 mg/L gentamicin, 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG), and 100 ppm Neutral Red dye. Briefly, after electroporation, an appropriate amount of the culture was spread on the Neutral Red plates, so that each plate had 300-500 colonies. The plates were incubated at 30 ℃ for three days and the colonies with the most intense red color were selected for further characterization with GC-MS.

Analysis of the fatty acid production with GC-MS

To characterize the fatty acid production and composition of each variant, the *in vivo* activity was determined according to the approach described previously with slight modification (Jing et al. 2011). Colonies were picked from Neutral Red plates, inoculated into 700 μL of LB medium supplemented with 100 mg/L carbenicillin, and cultured overnight at 30℃ at 250 rpm agitation rate. Next morning, 100 μL of the overnight culture was used to inoculate 2 mL M9 medium supplemented with 0.4% glucose, 100 mg/L carbenicillin and 0.1 mM ITPG in a 16-mL test tube. After culturing at 30℃ and 250 rpm for 48 hours, 1.5 mL of culture was used for fatty acid extraction. Following the addition of 50 μg heptanoic acid (7:0), 50 μg undecanoic acid (11:0), and 100 μg heptadecanoic acid (17:0) (Sigma-Aldrich, St. Louis, MO, USA) as internal standards, the mixture was acidified with 500 μL of 1 M HCl, and 4 mL chloroform-methanol (1:1 vol/vol) was used to recover the fatty acids from the culture. After vortexing for 10 min and centrifuging at 3000 × g for 4 min, the lower chloroform phase

was transferred onto anhydrous $MgSO_4$ column to remove trace amounts of water, and then evaporated under a stream of N_2 gas until the samples were concentrated to ~200 μL. These samples were subjected to GC-MS analysis. The fatty acid production of each acyl-ACP TE mutant was obtained by subtracting the fatty acids produced by *E. coli* expressing a control plasmid (pUCHisGm) that did not contain TE gene.

Results

Characterization of the variant acyl-ACP TE library

We initially evaluated the diversity of the acyl-ACP TE sequences in the evolved library that was constructed. Forty seven colonies were randomly picked from plates without any Neutral Red staining, and the acyl-ACP TE gene sequences were determined. Alignment of the nucleotide sequences of 47 variant TEs showed that they all have different sequences, suggesting that the assembly of the mutant library from the 30 degenerate oligos was random. Further analysis of these sequences indicated that 45 TE mutants contain nonsense mutations, such as premature stop codons, nucleotide insertions, or deletions. Only two TE mutants can be translated into full length proteins. This result suggests that the approach we used to generate the mutant library has very high probability to cause nonsense mutations (~95%).

Screening for TE activity on Neutral Red plates

All acyl-ACP TE variants were expressed in *E. coli* strain K27, which carries a mutation in the acyl-CoA synthetase gene (fadD) and is thus capable of accumulating free fatty acids. To discriminate the variants that have high TE activity from those having low activity, K27 cells expressing TE variants were screened on Neutral Red plates. Neutral Red is a pH indicator which changes color to red when pH drops below 6.8. The variants with higher TE activity will produce more free fatty acids, which will decrease the pH of the colonies and generate a more intense red color. After incubation at 30℃ for three days, red colonies of different intensities were visualized on the Neutral

Red plates (Figure 5.3). Approximately 98% of the colonies displayed light red color, and only about 2% of the colonies exhibited a more intensive red color.

Quantification of the fatty acid productivity of TE variants

To determine whether the fatty acid production is correlated with the color of colonies, 133 dark red colonies and 77 light red colonies were randomly picked from the Neutral Red plates for further analysis with GC-MS. For each colony expressing a TE mutant, the free fatty acids were extracted and analyzed, and the total fatty acid production was calculated. Of the 133 strains that expressed dark red colonies, 75% produced more than 600 μM of fatty acids, 50% produced more than 1000 μM fatty acids, and 25% produced more than 1200 μM. Only 25% of the dark red colonies produced less than 600 μM total fatty acids (Figure 5.4). In contrast, most of the light red colonies produced very small amounts of fatty acids (<100 μM) (Figure 5.4). The maximum fatty acid production of light red colonies is 264 μM, which is much lower than the production of most dark red colonies. Together, these results indicate that there is a strong correlation between the color of colonies and the total fatty acid productivity, indicating the validity of the Neutral Red screening protocol for identifying strains that express high fatty acid productivity.

Screening TE variants for higher fatty acid productivity

To identify TE variants from our variant library that express higher fatty acid productivity, 480 colonies were selected from the Neutral Red plates based upon the red-color colony phenotype, and their fatty acid productivity was analyzed with GC-MS. In parallel, the six plant acyl-ACP TEs that were initially used as guidance for the design of the variant library were also analyzed with GC-MS. The total fatty acid productions of CnFatB3, CvFatB1, CnFatB2, UaFatB1, CvFatB2, and CpFatB1 are 103, 243, 270, 352, 484, and 932 μM respectively (green bars in Figure 5.5). Among 480 colonies, 156 colonies produced more than 1000 μM total fatty acids, which is higher productivity than the highest of six parental acyl-ACP TEs. The fatty acid productivity of a subset of those

variants is shown in Figure 5.5. The highest fatty acid productivity observed is 1695 μM, which is about 80% higher than the productivity of CpFatB1. On a fatty acid weight basis, the highest productivity is 349 mg/L, which is 2.6 fold higher than that of CpFatB1.

Sequences of the mutants

A total of 192 acyl-ACP TE variants that expressed a fatty acid productivity of between 500 μM and 1700 μM were selected for high throughput sequencing; and 177 were successfully sequenced. Among these, 147 variants have the identical sequence, named TEGm162. This sequence was not discovered in the 47 randomly sequenced variants isolated from the non-screening plate, suggesting that the Neutral Red plate has strong screening capability for the variant TEGm162. With 147 analyses, the average fatty acid productivity of TEGm162 was 1173 ± 207 μM (Mean ± SD), which is about 26% higher than the productivity of CpFatB1 and 10-fold higher than the productivity of CnFatB3. Among the other sequenced variants, another three had the identical sequences, named TEGm204. Including TEGm162 and TEGm204, 27 distinct sequences were identified from 177 sequenced acyl-ACP TE variants. Among them, three variants had N-terminal truncations and one had a C-terminal truncation, but they still showed higher fatty acid productivity. The variant with the C-terminal truncation lacked 138 amino acids, and was not included in the following analysis. The protein sequences of 26 distinct variants and the 6 parental TEs were subjected to phylogenetic analysis. The phylogenetic tree indicates that the closest sequence to the six parental TEs is TEGm419, which has 78% sequence identity with CpFatB1 but shows broad substrate specificity (Figure 5.6). Some TE variants have similar sequences but display different substrate specificities, which can be useful for identifying residues that determine the specificity of acyl-ACP TE.

The substrate specificities of TE mutants

Although the Neutral Red plates screening strategy can only distinguish colonies on the basis of fatty acid productivity, because the six input acyl-ACP TEs showed different substrate specificities, we were interested to evaluate how this characteristic evolved in the variants that showed enhanced catalytic efficiency. The fatty acid profiles of the 26 distinct variants and 6 parental TEs are shown in Figure 5.6. Those variants exhibit various substrate specificities: 16 variants produced mainly 14/16-carbon fatty acids; 4 produced mainly 8-carbon fatty acid; and 6 produced evenly distributed 8- to 16-carbon fatty acids. The diverse sequences and substrate specificities of those mutants provide a good opportunity to study the sequence-function relationship. In the phylogenetic tree (Figure 5.6), TEGm205 and TEGm258 have very similar protein sequences, but they display different substrate specificities (Figure 5.6). Comparing the sequences of TEGm205 and TEGm258 identifies the 9 different residues between these two proteins, which occur at the N-terminal one-third of the enzyme. Among these, residue 92 (numbered based on sequence of TEGm258) has been shown to affect the substrate specificity in the chapter four. Another three different residues are in the region that includes residues 55 through 67, which is also proposed to affect the substrate specificity in the chapter four. Similarly, comparing the sequences of TEGm202 and TEGm157, as well as TEGm201 and TEGm245, also leads to the identification of different residues in this region, which may be responsible for the different substrate specificities of these mutants. All these data provide further support for our previous hypothesis that the region that includes residues 55 through 67 is critical for the substrate specificity of acyl-ACP TE.

Discussion

In this study, a directed evolution approach has been developed to increase the fatty acid productivity of acyl-ACP TE. Based on the results, some optimization can be made to further improve this approach. Generation of a mutant library is the first critical step for directed evolution. In this study, a designed library was constructed by

assembling 30 degenerate primers. Among 307 amino acid positions, 65 positions identified from multiple sequence alignment of six plant acyl-ACP TEs were chosen to incorporate variable amino acids. The theoretical size of this mutant library is 10^{38}. However, the actual size is much smaller. For this 921 bp gene, 1 microgram of DNA contains 10^{12} molecules. Considering the bias and enrichment effect of PCR, 1 microgram of PCR-amplified DNA may contain 10^{6}-10^{7} variants. The number of colonies we can screen on the Neutral Red plate is at the level of 10^{4}, and the number of samples we can analyze with GC-MS is several hundreds. This very large variant library was designed at the beginning in the expectation that a library of larger size would give us better chance to find improvements in TE activity. However, with limited screening throughput, the large size library does not enhance the chance, because it is impossible to screen even a small portion of such a large library. Many examples of directed evolution studies show successful improvements can be achieved with small libraries, suggesting that smaller and higher quality mutant libraries can reduce the screening effort without compromising the chance of finding significant improvements (Turner 2009). Therefore, in future directed evolution studies of acyl-ACP TE, more effort should be made to generate a high quality variant library.

Assembling the variant library from small oligos results in the incorporation of many nonsense mutations, such as premature stop codons, nucleotide insertions, and deletions. To eliminate the nonsense mutations, TE variants were fused at the C-terminus to a gentamicin resistant gene. In this study, 2.5 mg/L gentamicin was supplemented into the screening plate, but this failed to eliminate all nonsense mutants. When randomly sequencing mutants from the non-screening plate, we frequently picked out mutants with premature stop codons, suggesting the concentration of gentamicin was not sufficient to inhibit the cells that do not express the gentamicin resistant gene. In the future, the concentration of gentamicin in the screening plate needs to be further optimized.

Despite the need for future improvements, this study represents the development of first directed evolution protocol and result in the increased catalytic activity and more

insights on the relationship between sequence and substrate specificity of acyl-ACP TE. In this study, TE variants were expressed in *E coli* strain K27, which is capable of accumulating free fatty acids. Using Neutral Red as a pH indicator, the K27 cells expressing highly active TE variants can be discriminated from those expressing non-active mutants. We demonstrated that the recovery of dark red colored colonies is correlated with high fatty acid production. The exact fatty acid productivity of those active variants was further analyzed with GC-MS. In just one round of directed evolution, several TE mutants with significantly increased fatty acid productivity were identified from the designed mutant library. Specifically, TEGm162 was recovered with very high frequency. The fatty acid production of TEGm162 is 30% more than the most active parental TE CpFatB1 and 10-fold more than the least active parental TE CnFatB3. In this study, we also demonstrate that directed evolution can be used to study the sequence-function relationship of acyl-ACP TEs, in terms of substrate specificity. Comparison of the sequences and fatty acid profiles of those mutants suggest that the region from position 55 to 67 is important for the determination of substrate specificity of acyl-ACP TE. This is consistent with the hypothesis supported by structurally directed studies described in the previous chapter. Acyl-ACP TEs have drawn intensive attention in recent efforts to produce fatty acids as biofuels or biochemicals, because they control the hydrolysis reaction of acyl-ACPs and thus the termination of the acyl-chain elongation in the FAS pathway (Lennen et al. 2012). This directed evolution approach, as well as the sequence-function relationship knowledge obtained in this study, will be useful for improving the acyl-ACP TE for the application in metabolic engineering.

Acknowledgement

This work was supported by the U.S. National Science Foundation through its Engineering Research Center Program (Award No. EEC-0813570). The authors thank M. Ann D.N. Perera and Zhihong Song of the W.M. Keck Metabolomics Research Laboratory at Iowa State University for assistance with fatty acid analysis. We thank Sara Pederson for helping with the mutant library generation and fatty acid analysis. We

thank Le Zhao for picking the colonies and making glycerol stocks. We thank Derek Loneman for assistance with fatty acid extraction.

References

Chen, X., Zaro, J. L. and Shen, W. C. (2012). Fusion protein linkers: Property, design and functionality. Adv Drug Deliv Rev.

Cobb, R. E., Sun, N. and Zhao, H. M. (2013). Directed evolution as a powerful synthetic biology tool. Methods 60(1): 81-90.

Dougherty, M. J. and Arnold, F. H. (2009). Directed evolution: new parts and optimized function. Current Opinion in Biotechnology 20(4): 486-491.

Durrett, T. P., Benning, C. and Ohlrogge, J. (2008). Plant triacylglycerols as feedstocks for the production of biofuels. Plant Journal 54(4): 593-607.

Dyer, J. M., Stymne, S., Green, A. G. and Carlsson, A. S. (2008). High-value oils from plants. Plant Journal 54(4): 640-655.

Facciotti, M. T. and Yuan, L. (1998). Molecular dissection of the plant acyl-acyl carrier protein thioesterases. Fett-Lipid 100(4-5): 167-172.

Handke, P., Lynch, S. A. and Gill, R. T. (2011). Application and engineering of fatty acid biosynthesis in Escherichia coli for advanced fuels and chemicals. Metabolic Engineering 13(1): 28-37.

Heath, R. J. and Rock, C. O. (1996). Inhibition of beta-ketoacyl-acyl carrier protein synthase III (FabH) by acyl-acyl carrier protein in Escherichia coli. Journal of Biological Chemistry 271(18): 10996-11000.

Jiang, P. and Cronan, J. E. (1994). Inhibition of Fatty-Acid Synthesis in Escherichia-Coli in the Absence of Phospholipid-Synthesis and Release of Inhibition by Thioesterase Action. Journal of Bacteriology 176(10): 2814-2821.

Jing, F., Cantu, D. C., Tvaruzkova, J., Chipman, J. P., Nikolau, B. J., Yandeau-Nelson, M. D. and Reilly, P. J. (2011). Phylogenetic and experimental characterization of an acyl-ACP thioesterase family reveals significant diversity in enzymatic specificity and activity. BMC Biochemistry 12: 44.

Lennen, R. M. and Pfleger, B. F. (2012). Engineering Escherichia coli to synthesize free fatty acids. Trends in Biotechnology 30(12): 659-667.

Leonard, J. M., Knapp, S. J. and Slabaugh, M. B. (1998). A Cuphea beta-ketoacyl-ACP synthase shifts the synthesis of fatty acids towards shorter chains in Arabidopsis seeds expressing Cuphea FatB thioesterases. Plant Journal 13(5): 621-628.

Liu, X., Sheng, J. and Curtiss, R., 3rd (2011). Fatty acid production in genetically modified cyanobacteria. Proceedings of the National Academy of Sciences of the United States of America 108(17): 6899-6904.

Magnuson, K., Jackowski, S., Rock, C. O. and Cronan, J. E., Jr. (1993). Regulation of fatty acid biosynthesis in Escherichia coli. Microbiological Reviews 57(3): 522-542.

Mayer, K. M. and Shanklin, J. (2007). Identification of amino acid residues involved in substrate specificity of plant acyl-ACP thioesterases using a bioinformatics-guided approach. BMC Plant Biology 7: 1.

Nair, N. U. and Zhao, H. M. (2008). Evolution in reverse: Engineering a D-xylose-specific xylose reductase. Chembiochem 9(8): 1213-1215.

Nikolau, B. J., Perera, M. A., Brachova, L. and Shanks, B. (2008). Platform biochemicals for a biorenewable chemical industry. Plant Journal 54(4): 536-545.

Ohlrogge, J. B. (1994). Design of New Plant Products: Engineering of Fatty Acid Metabolism. Plant Physiology 104(3): 821-826.

Pollard, M. R., Anderson, L., Fan, C., Hawkins, D. J. and Davies, H. M. (1991). A specific acyl-ACP thioesterase implicated in medium-chain fatty acid production in immature cotyledons of Umbellularia californica. Archives of Biochemistry and Biophysics 284(2): 306-312.

Ranganathan, S., Tee, T. W., Chowdhury, A., Zomorrodi, A. R., Yoon, J. M., Fu, Y., Shanks, J. V. and Maranas, C. D. (2012). An integrated computational and experimental study for overproducing fatty acids in Escherichia coli. Metabolic Engineering 14(6): 687-704.

Richardson, S. M., Wheelan, S. J., Yarrington, R. M. and Boeke, J. D. (2006). GeneDesign: Rapid, automated design of multikilobase synthetic genes. Genome Research 16(4): 550-556.

Steen, E. J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., Del Cardayre, S. B. and Keasling, J. D. (2010). Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. Nature 463(7280): 559-562.

Thelen, J. J. and Ohlrogge, J. B. (2002). Metabolic engineering of fatty acid biosynthesis in plants. Metabolic Engineering 4(1): 12-21.

Turner, N. J. (2009). Directed evolution drives the next generation of biocatalysts. Nature Chemical Biology 5(8): 568-574.

Voelker, T. A. and Davies, H. M. (1994). Alteration of the specificity and regulation of fatty acid synthesis of Escherichia coli by expression of a plant medium-chain acyl-acyl carrier protein thioesterase. Journal of Bacteriology 176(23): 7320-7327.

Voelker, T. A., Worrell, A. C., Anderson, L., Bleibaum, J., Fan, C., Hawkins, D. J., Radke, S. E. and Davies, H. M. (1992). Fatty acid biosynthesis redirected to medium chains in transgenic oilseed plants. Science 257(5066): 72-74.

Youngquist, J. T., Lennen, R. M., Ranatunga, D. R., Bothfeld, W. H., Marner, W. D. and Pfleger, B. F. (2012). Kinetic modeling of free fatty acid production in Escherichia coli based on continuous cultivation of a plasmid free strain. Biotechnology and Bioengineering 109(6): 1518-1527.

Zha, W. J., Rubin-Pitel, S. B. and Zhao, H. M. (2008). Exploiting genetic diversity by directed evolution: molecular breeding of type III polyketide synthases improves productivity. Molecular Biosystems 4(3): 246-248.

Zhang, F., Carothers, J. M. and Keasling, J. D. (2012). Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. Nature Biotechnology 30(4): 354-359.

Zhang, X. J., Li, M., Agrawal, A. and San, K. Y. (2011). Efficient free fatty acid production in Escherichia coli using plant acyl-ACP thioesterases. Metabolic Engineering 13(6): 713-722.

Tables

Table 5.1 Sequences of designed oligos for variant library construction

oligo name	sequences (5'->3')
M1-1	TCATGGATCCGGAAAAACAGTGGACCMTGTTWGACYGKAAACCGARACGTCCGGACATGC
M1-2	CGAHACGACCCAGACCGAAAGVGTCCAYCAGCATGTCCGGACGTYTCGGT
M1-3	CTTTCGGTCTGGGTCGTDTCGTTMAGRACGGTCTGGTTTTCCGTCAGARC
M1-4	GATTTCGTAAGAACGGATAGAGAAGYTCTGACGGAAAACCAGACCGT
M1-5	TTCTCTATCCGTTCTTACGAAATCGGTGYTGACCGTACCGCTTCTRTCGAAACC
M1-6	GTGGTTGAKCGHGGTTTCCTGMAMGTGGTTCATCASGGTTTCGAYAGAAGCGGTACG
M1-7	CAGGAAACCDCGMTCAACCACKKCAAATCTNYCGGTCTGMTGRACGACGGTTTC
M1-8	CAGGTYACGTTTGBACATTTCCGGGGTACGACCGAAACCGTCGTYCAKCAGACC
M1-9	CCGGAAATGTVCAAACGTRACCTGATCTGGGTTNTCRCCAAAATGCAGRTCATGRTCRACC
M1-10	GTCACCCCAGGTCGGGTAACGGTYGAYCATGAYCTGCATTTTGGYGA
M2-1	CGTTACCCGACCTGGGGTGACACCNTTGAGNTCAACACCTGGNTTTCTMAATCTGGTAAAA
M2-2	CAGCCAGTCACKASSCATACCGWTTTTACCAGATTKAGAAANCCAGGTGTTG
M2-3	GTATGSSTMGTGACTGGCTGRTCHSTGACTGCMAMACCGGTGAAAYCMTCVTSCGT
M2-4	GGTTTTCTKGTTCATCATARCSYAAACAGAGGTAGCACGSABGAKGRTTTCACCGGT
M2-5	TGTTTRSGYTATGATGAACMAGAAAACCCGTARANTCTCTARANTCCCGGAAGAAGTT
M2-6	GAGTCGANGAAGTRCGGASCGANCTCCYGACGAACTTCTTCCGGGANTYTAGAGA
M2-7	TCGSTCCGYACTTCNTCGACTCTSCGSCGGYTNTCGAAGACRACGACSGTAAACTGC
M2-8	GATAGAGTCASCGSTTTTCWCGTCGANTTTCBGCAGTTTACSGTCGTYGTCTTCGA
M2-9	CGACGWGAAAASCGSTGACTCTATCARAARAGGTCTGACCCCGCGTTGGARCGAC
M2-10	AACGTGCTGGTTGANGTCGANGTCGYTCCAACGCGGGGTCAGA
M3-1	GACNTCGACNTCAACCAGCACGTTARCAACGYGAAATACNTCGGTTGGNTTCTG
M3-2	AGAAYTTCGRTCGGCAYAGATTCCAGAANCCAACCGANGTATTTCRCGTT

Table 5.1 continued

M3-3	GAATCTRTGCCGAYCGAARTTCTGGAAACCCAKGAGNTCKSCTCTNTG
M3-4	ACCGCATTCACGACGGTATTCCAGGSTCANAGAGSMGANCTCMTGGG
M3-5	CTGGAATACCGTCGTGAATGCGGTCGTGACTCTGTTSTGSAATCTSTTACCTCTRTG
M3-6	GANACGGYCAYCCWCTTYAGACGGGYYCAYAGAGGTAASAGATTSCASAACAGA
M3-7	CCGTCTRAAGWGGRTGRCCGTNTCSAGTRCCRGCACCTGSTGCGTCTGG
M3-8	TTCGGTACGACCTTTCAYGAWSTCAGYACCGTCTTCCAGACGCASCAGGTGCYG
M3-9	GASWTCRTGAAAGGTCGTACCGAATGGCGTCCGAAAAACGCTGGTACCAACG
M3-10	AGGTGAATTCTTGGTTTTACCGGTAGAGATAGCACCGTTGGTACCAGCGTTTTTCGG

Figures

Figure 5.1 Multiple sequence alignment of six parental acyl-ACP TEs

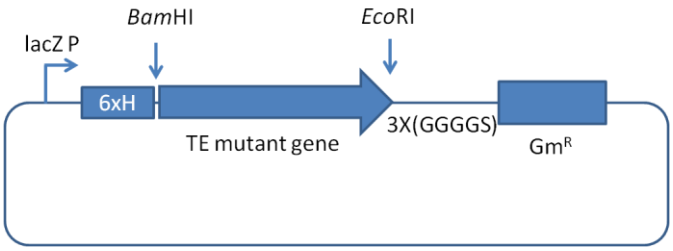


Figure 5.2 Schematic diagram of TE variant in pUCHisGm vector

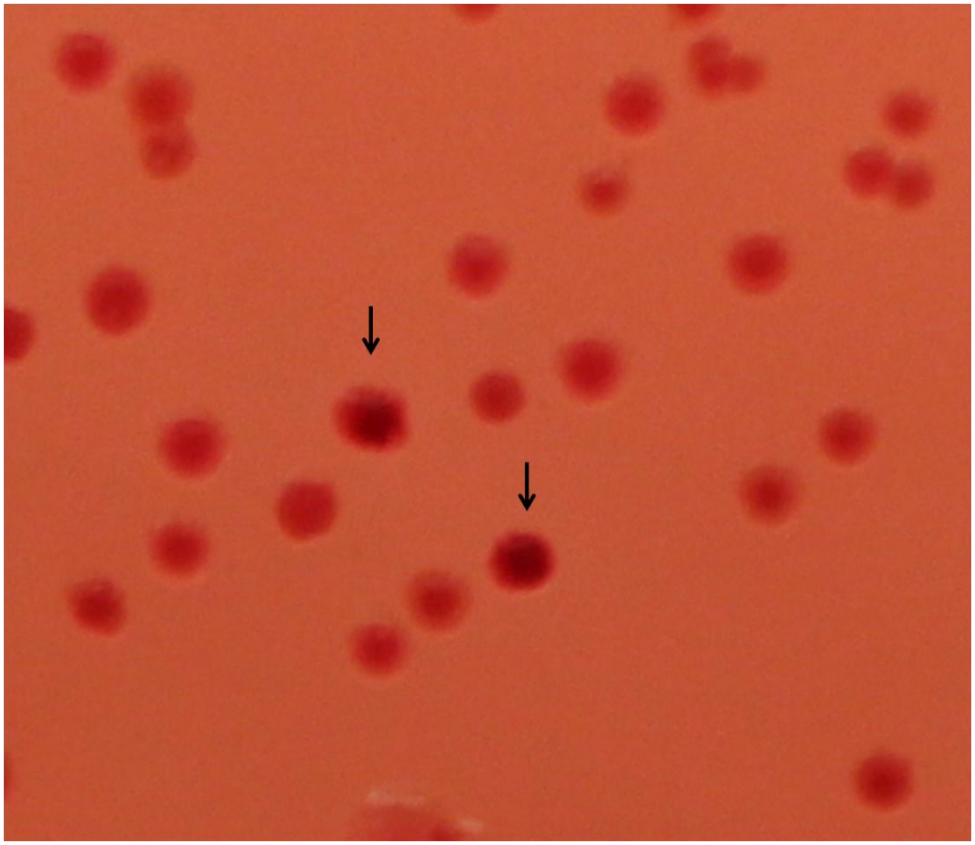


Figure 5.3 Colonies expressing TE variants on the Neutral Red Plate. The colonies displaying more intensive red color are indicated by arrows.

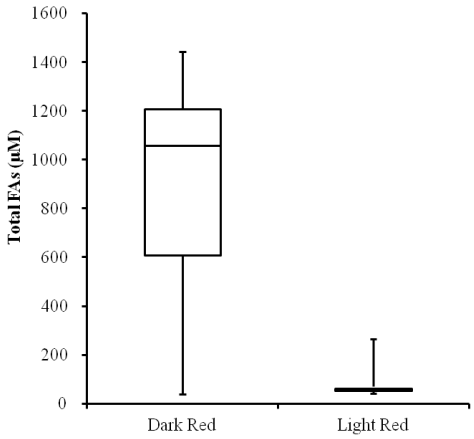


Figure 5.4 Box-and-whisker plot of fatty acid production of dark red and light red colonies. 177 dark red colonies and 77 light red colonies were analyzed.

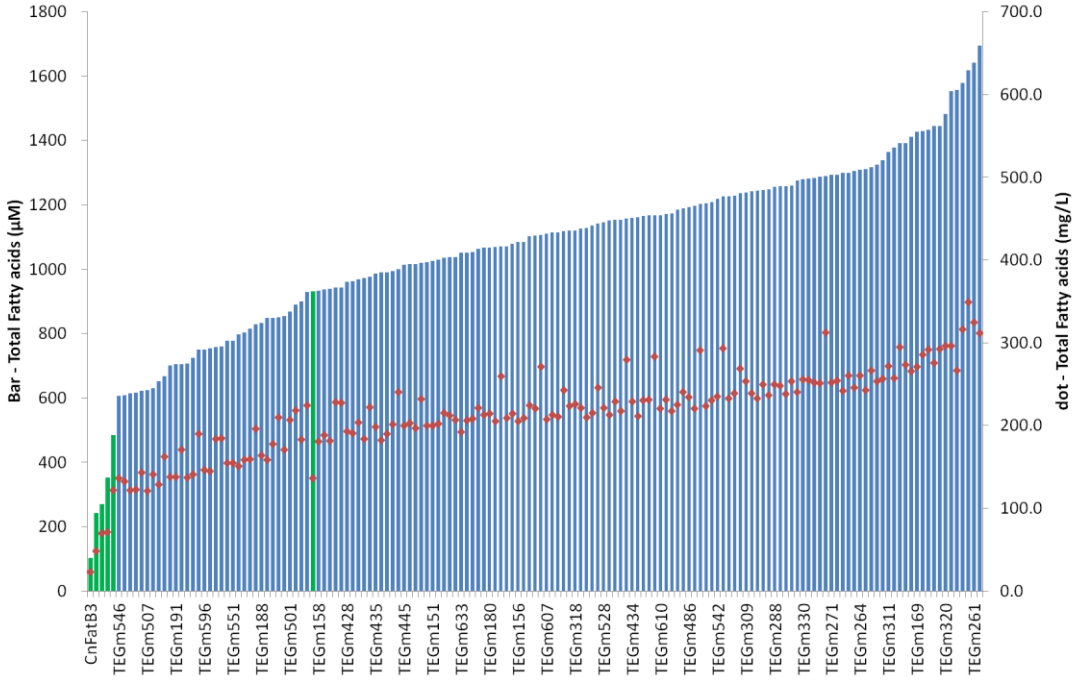
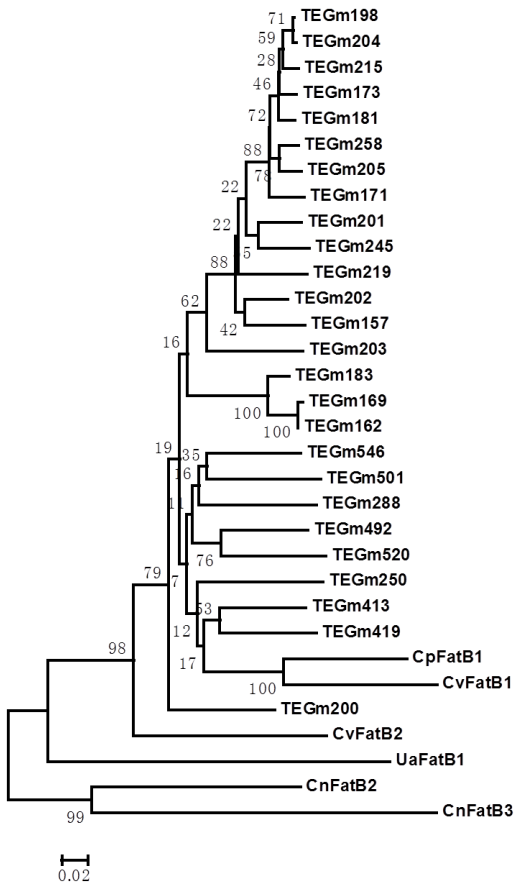


Figure 5.5 Fatty acid productivities of parental TEs (green bars) and TE variants (blue bars). Bars represent the total fatty acids in µM and dots represent the total fatty acids in mg/L.

A



B

mutant names	Total FAs(uM)	Total FAs (mg/L)	mol percentage (%)											
			4:0	6:0	8:0	10:0	10:1	12:0	12:1	14:0	14:1	16:0	16:1	18:1
TEGm198	705	170	0.1	0.3	1.3	0.1	0.0	1.4	0.1	36.4	0.7	23.0	36.6	0.0
TEGm204	1203	291	0.1	0.2	1.3	0.1	0.0	1.6	0.3	38.0	1.5	19.8	33.6	3.4
TEGm215	977	222	1.6	1.0	9.0	0.7	0.1	5.6	1.7	30.6	6.5	15.2	24.3	3.7
TEGm173	1226	293	0.3	0.3	2.8	0.4	0.0	3.5	0.5	35.9	3.1	12.0	36.2	4.9
TEGm181	829	196	0.4	0.5	3.7	0.7	0.0	4.8	0.5	35.5	4.6	15.1	29.2	5.0
TEGm258	1558	266	4.4	6.4	50.4	6.4	1.1	4.7	1.8	9.1	1.3	10.2	1.9	2.3
TEGm205	891	218	0.0	0.1	0.3	0.0	0.0	0.7	0.0	35.9	1.1	23.9	34.1	3.9
TEGm171	1118	242	1.3	1.4	17.6	0.8	0.1	6.4	3.1	26.2	5.9	12.1	23.0	2.0
TEGm201	594	148	0.1	0.2	1.3	0.1	0.0	1.2	0.1	19.3	1.3	28.4	41.1	7.0
TEGm245	1106	270	0.1	0.0	1.0	0.0	0.0	0.7	0.0	32.1	0.0	27.7	38.5	0.0
TEGm219	528	123	0.1	0.5	2.4	0.3	0.0	4.4	0.2	54.7	3.8	16.1	17.5	0.0
TEGm202	1001	240	0.4	0.4	2.8	0.3	0.0	4.0	0.8	30.1	2.9	17.9	36.4	4.0
TEGm157	598	137	0.4	1.0	12.0	0.7	0.0	4.0	0.5	30.9	4.6	19.1	19.2	7.6
TEGm203	526	127	0.1	0.1	0.6	0.1	0.0	2.4	0.0	45.1	2.2	24.1	19.6	5.7
TEGm183	1236	269	0.4	0.5	12.7	4.2	0.8	9.2	6.6	23.2	11.7	10.0	18.4	2.4
TEGm169	1427	271	0.4	4.2	33.7	4.6	0.6	11.7	2.8	18.8	14.6	3.8	1.9	2.7
TEGm162	1695	311	0.5	4.9	38.5	5.3	0.5	11.7	2.2	19.4	11.6	3.1	0.6	1.5
TEGm546	607	136	0.0	0.3	1.3	5.2	0.6	21.1	7.1	22.1	11.7	14.9	12.2	3.4
TEGm501	868	206	0.0	0.1	0.4	0.4	0.0	9.1	1.0	38.7	4.9	17.1	23.7	4.6
TEGm288	1256	249	0.2	0.3	30.9	5.5	2.2	9.8	8.5	12.9	7.1	12.1	8.3	2.2
TEGm492	604	146	0.2	0.2	2.5	1.5	0.0	8.6	0.0	21.9	2.8	25.7	24.0	12.5
TEGm520	1021	232	0.0	0.2	2.4	3.1	1.5	8.8	10.3	24.0	15.0	11.4	20.6	2.8
TEGm250	1618	349	1.6	1.7	20.0	4.3	0.7	3.8	2.0	19.8	3.5	19.1	17.1	6.2
TEGm413	1142	245	0.1	0.3	14.8	6.6	1.4	13.6	7.6	14.2	5.9	14.7	16.9	3.9
TEGm419	1392	295	0.1	0.2	13.2	7.6	2.0	13.7	12.2	15.2	8.0	8.9	15.2	3.6
CpFatB1	932	136	0.01	1.51	93.46	2.18	0.44	0.87	0.21	0.00	0.10	1.21	0.00	0.00
CvFatB1	243	48	0.00	1.22	24.16	21.59	1.48	4.84	3.89	21.10	0.21	10.17	5.60	5.74
TEGm200	944	227	0.3	0.0	1.8	0.4	0.1	4.7	3.7	19.2	8.3	16.6	41.5	3.5
CvFatB2	484	122	0.00	0.00	0.67	0.56	0.15	0.31	0.24	27.85	0.00	39.95	11.22	19.06
UaFatB1	352	71	0.07	0.31	21.79	19.72	0.93	3.56	1.95	32.23	0.27	8.37	6.27	4.54
CnFatB2	270	70	0.00	0.00	0.14	0.00	0.00	0.00	0.00	13.77	0.00	52.57	7.45	26.08
CnFatB3	103	23	0.00	0.12	6.02	0.00	0.00	34.29	2.28	26.17	5.71	12.17	3.96	9.27

Figure 5.6 Phylogenetic analysis (A) and fatty acid profiles (B) of TE variants and parents. The phylogenetic tree was inferred using the Minimum Evolution method. The bootstrap consensus tree inferred from 250 replicates is taken to represent the evolutionary history.

Figure 5.7 sequence alignment of TE variants TEGm258 and TEGm205. The region proposed in the previous chapter to affect the substrate specificity of acyl-ACP TE is indicated by a line. The dot indicated residue has been proved to affect specificity in the chapter four. The residues proposed here to be involved in determining the different substrate specificities are indicated by squares.

CHAPTER 6 – CONCLUSIONS

Prior characterizations of acyl-ACP TEs have primarily focused on plant sourced TEs and their substrate specificity on acyl-ACPs with different acyl chain lengths. In this study, I explored the functional diversity of acyl-ACP TEs sourced from a wider taxonomic clades. A total of 31 TE enzymes sourced from both plants and bacteria were functionally characterized. The results reveal that bacterial acyl-ACP TEs provide access to additional functional diversity, both relative to acyl chain length specificity (e.g., shorter acyl chains, as short as four carbon atoms), as well as acyl chains that contain additional chemical functionalities (e.g., unsaturated acyl chains and acyl chains containing a carbonyl group). This additional functional diversity in acyl-ACP TEs can potentially be used to diversify the fatty acid biosynthesis pathway to produce biorenewable chemicals.

Plant acyl-ACP TEs were previously proposed to utilize a papain-like protease catalytic triad consisting of Cys, His, and Asn residues. In this study, multiple sequence alignment of plant and bacterial TEs, and structure-modeling of CvFatB2 TE suggested that the previously proposed Cys348 residue is unlikely to be a catalytic residue. Instead, residues Asp309 and Glu347, in addition to previously proposed residues Asn311 and His313 (numbers are based on CvFatB2 sequence), are proposed to be involved in the catalysis of acyl-ACP TEs. *In vivo* activities of site-directed mutants tested and supported this hypothesis. By comparing the bacterial TE structure (2OWN) and *Pseudomonas* 4HBT structure (1BVQ), a two-step catalytic mechanism for plant and bacterial acyl-ACP TEs is proposed.

A domain shuffling approach was used to identify the region(s) that determine the substrate specificity of two acyl-ACP TEs (CvFatB1 and CvFatB2). The CvFatB1 and CvFatB2 proteins were dissected into six equal-length domains (Fragment I-VI), and shuffling of these domains was used to generate a series of chimeric proteins. Comparison of the substrate specificities of the chimeric TEs indicate that Fragment III is the most import region for determining substrate specificity, and Fragments I, II and

IV affect specificity to a lesser extent, and Fragments V and VI are not involved in specificity determination. Structural modeling of CvFatB2 revealed a hotdog structure. The N-terminal hotdog domain comprises mostly Fragments II and III, and is therefore the structural basis for the substrate specificity. Based on a multiple sequence alignment of 21 functionally characterized plant acyl-ACP TEs, nine residues were predicted to affect the substrate specificity. Site-directed mutagenesis analysis demonstrated that six residues play critical roles in determining the substrate specificity, including V194 in Fragment II, V217, N223, R226, and R227 in Fragment III, and I268 in Fragment IV. Another three residues, L257, I260, and L289, were shown to impact the catalytic activity of acyl-ACP TE, because they are in two proposed ACP binding motifs. Analysis of a crystal structure of acyl-ACP TE from *Lactobacillus plantarum* (2OWN) and comparing that with the experimental data in this study lead to a hypothesis that plant acyl-ACP TE has a similar cavity in the N-terminal hotdog domain between the central α-helix and the anti-parallel β-sheets, which determines the substrate specificity.

A directed evolution approach was developed to improve the *in vivo* fatty acid productivity of acyl-ACP TE. A variant library was designed and screened on Neutral Red plates, resulting in the recovery of TE variants with increased fatty acid productivity and more insight into the sequence-specificity relationship of acyl-ACP TEs.

The significant diversity of the enzymatic activity and specificity of the natural acyl-ACP TEs demonstrate the plasticity of this enzyme, which is a characteristic that will enable further protein engineering of this biocatalyst. The better understanding of the catalytic residues and catalytic mechanism will allow us to engineer acyl-ACP TEs with higher catalytic efficiency. The identification of the region and residues that determine the substrate specificity of acyl-ACP TEs will benefit the engineering of this biocatalyst with desired substrate specificity. The developed directed evolution approach provides a powerful tool for tailoring this biocatalyst.